Magnitude Processing of Written Number Words is Influenced by Task, Rather Than Notation

Becky Wong[a], Rebecca Bull[a], and Daniel Ansari[b, a]

[a]National Institute of Education, Nanyang Technological University, Singapore

[b]Numerical Cognition Laboratory, Department of Psychology, Brain and Mind Institute, The University of Western Ontario, Canada

Full article available at: https://doi.org/10.1016/j.actpsy.2018.09.010

Word Count: 9434 (Excludes abstract, appendices, and references)

Corresponding author and details: Becky Wong, National Institute of Education, NIE-3-02-02, 1 Nanyang Walk, Singapore 637 616, Singapore. E-mail: nie15661@e.ntu.edu.sg

Abstract

The extent to which task and notation influence the processing of numerical magnitude is under theoretical and empirical debate. To date, behavioural studies have yielded a mixed body of evidence. Using the case of written number words in English and Chinese, we re-examined this issue. Thirty-nine bilingual participants who showed a balanced profile of language dominance in English and Chinese completed three tasks of numerical processing (Magnitude Comparison, Numerical Matching, and Language Matching) with pure English, pure Chinese, and mixed notation number words. We conducted frequentist and Bayesian statistics on the data. Magnitude processing, as indexed by the numerical distance effect (NDE), was found to be dependent on task. Specifically, the NDE occurred in all notation conditions in the Magnitude Comparison Task and mixed notation trials in the Numerical Matching Task only. However, the data indicated that magnitude processing was independent of notation. Task and notation had an interactive influence on overall speed of processing, where participants responded to Chinese number words significantly faster than other notations for the Magnitude Comparison and Numerical Matching Tasks only. Finally, Bayesian analyses indicated that task and notation do not interact to affect magnitude processing. Specifically, the Bayes Factor and posterior model probabilities of the Bayesian ANOVA yielded strongest support for the model with three main effects (Task, Notation, Numerical Distance) and two two-way interactions (Task x Numerical Distance, Task x Notation). These findings highlight the critical role of task in numerical magnitude processing, provide support for a notation-independent account of magnitude processing, and suggest that linguistic / orthographic factors, combined with task, may interact to affect overall speed of processing.

(267 words)

Keywords: Bilingualism; Numerical Cognition; Task-Dependence; Notation-Independence; Magnitude Processing

Highlights

- Numerical magnitude processing is dependent on task, but independent of notation
- Task and notation have an interactive influence on overall reaction time
- Task and notation do not interact to influence magnitude processing

Suggested Classification Code: 2340 Cognitive Processes

Magnitude Processing of Written Number Words is Influenced by Task, Rather Than Notation

1. Introduction

Humans are able to recognise numbers in multiple notations (e.g., four, 4, ****) and process their magnitudes (e.g., judging that 7 is larger than 2). However, the extent to which task and notation conditions influence the processing of numerical magnitude is under debate. Additionally, existing studies have not yet directly addressed whether task and notation have an interactive influence on numerical magnitude processing or not. From a cognitive perspective, addressing these questions will help determine the conditions under which magnitude processing occurs, and the extent to which magnitude processing can be considered abstract or not.

1.1 The Influence of Task on Numerical Magnitude Processing

Various tasks are commonly used to investigate numerical magnitude processing. One group of tasks require participants to intentionally access and process the quantity of the numerical stimuli to successfully complete the task. These include the larger / smaller than X task (e.g., Cao, Li, & Li, 2010[1]; Cohen Kadosh, 2008; Lukas, Krinzinger, Koch, & Willmes, 2014; Pinel, Dehaene, Rivere, & LeBihan, 2001; Szucs & Csepe, 2005), the Magnitude Comparison Task (e.g., Cohen, Warren, & Blanc-Goldhammer, 2013; Goldfarb, Henik, Rubinsten, Block-David, & Gertner, 2011), and the numerical size condition in the Numerical Stroop Task (e.g., Ganor-Stern & Tzelgov, 2008; Ganor-Stern & Tzelgov, 2011; Girelli, Lucangeli, & Butterworth, 2000; Ito & Hatta, 2003; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002; Tzelgov, Meyer, & Henik, 1992). In these tasks, participants are instructed to respond

[1] All studies reviewed in the Introduction presented stimuli in the visual (written) modality, unless otherwise specified.

based on numerical magnitude, such as to select one out of two numbers that is larger in quantity, or to respond to whether each stimulus number is larger or smaller than a referent number.

Other tasks do not explicitly require participants to access and process the quantity of the numerical stimuli (henceforth referred to as ‘non-intentional tasks’). These include the Numerical Matching (Goldfarb et al., 2011; Sasanguie & Reynvoet, 2014; Verguts & Van Opstal, 2005; Van Opstal & Verguts, 2011) and Physical Matching tasks (Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008), and the physical size condition in the Numerical Stroop Task (Cohen Kadosh, Henik, & Rubinsten, 2008; Ito & Hatta, 2003; Ganor-Stern & Tzelgov, 2008; Ganor-Stern & Tzelgov, 2011; Rubinsten et al., 2002). In these tasks, numerical magnitude is not directly relevant. In order to follow the instructions, participants do not need to activate the numerical magnitudes associated with the numerical symbols they are asked to process. Instead, participants respond to various properties of the numbers, such as whether the numerical values or the language of the numbers are the same or different, or to select the numerical symbols or written number word that is larger in physical size.

In the tasks described above, common indicators of magnitude processing include the congruity effect in the physical size condition of Numerical Stroop Tasks (e.g., Cohen Kadosh et al., 2008; Girelli et al., 2000; Schwarz & Ischebeck, 2003) and the numerical distance effect (NDE). Our study focuses on the NDE as an indicator of magnitude processing.

The NDE refers to the well-known finding that participants are typically faster and more accurate when judging between numerosities of a large numerical distance (e.g., 2 9, Distance = 7) than numerosities of a small numerical distance (e.g., 3 4, Distance = 1). Exactly which stage of cognitive processing is the locus of the NDE remains under debate. According to the dominant view, the NDE arises from the stage of magnitude representation (e.g., Dehaene, 2003; Mazzocco, Feigenson, & Halberda, 2011) where it is thought that magnitude representations have an imprecise activation pattern, resulting in more representational overlap for numerosities of a small numerical distance than those with a large distance. Other alternative views are that the NDE arises from the stage of response selection (van Opstal, Gevers, de Moor, & Verguts, 2008; Verguts, Fias, & Stevens, 2005) or a more general mechanism that can be used to comparisons regarding time, space and quantity (Cohen Kadosh et al., 2008; also see Krajcsi, Lengyel, & Kojouharova, 2018).

Many studies have investigated either intentional or non-intentional tasks of magnitude processing. With respect to intentional tasks, studies have consistently reported the NDE in the larger than / smaller than X task for number words in Turkish (Lukas et al., 2014), Chinese (Cao et al., 2010), Hungarian (Szucs & Csepe, 2005), English (Pinel et al., 2001), Hebrew (Cohen Kadosh, 2008), and for Arabic numbers (e.g., Lukas et al., 2014; Pinel et al., 2001). Studies have also reported the NDE in the Magnitude Comparison task for number words in English (Cohen et al., 2013), Chinese (Campbell, Kanz, & Xue, 1999), and for Arabic numbers (e.g., Campbell et al., 1999; Duncan & McFarland, 1980; Goldfarb et al., 2011; Moyer & Landauer, 1967). Taken together, these results indicate that participants process numerical magnitude as instructed under intentional task conditions.

The NDE is less consistently reported in non-intentional tasks of magnitude processing. For example, in the Numerical Matching Task, studies have reported the NDE for mixed notation (Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008; Van Opstal & Verguts, 2011; Verguts & Van Opstal, 2005), but not mixed modality (i.e., auditory vs written) numbers (Cohen et al., 2013; Sasanguie & Reynvoet, 2014). Using this same task, other studies have reported the

NDE for written number words in English (Dehaene & Akhavein, 1995) and Indian (Ganor-Stern & Tzelgov, 2008), but not Arabic numbers (Goldfarb et al., 2011; Wong & Szucs, 2013). With respect to the Physical Matching Task, Dehaene and Akhavein (1995) reported the NDE for English number words and Arabic numbers, but not mixed English-Arabic notation, whereas Ganor-Stern and Tzelgov (2008) did not find the NDE for Indian, Arabic, or mixed Indian-Arabic notation.

Few studies have investigated both intentional and non-intentional tasks of magnitude processing. Among these studies, some suggest that numerical magnitude processing is independent of task. Using a Numerical Stroop Task, Ganor-Stern and Tzelgov (2008) reported a NDE for Indian and Arabic numbers for the numerical size condition, and a size congruity effect for the physical size condition. Pina, Castillo, Cohen Kadosh, and Fuentes (2015) reported similar findings for Arabic numbers only. Moreover, Tzelgov et al. (1992) presented participants with a referent number to memorize, and then showed them one number at a time. In the intentional task condition, participants judged whether each trial was larger or smaller than the referent number in numerical magnitude. In the non-intentional task condition, participants judged whether each trial was larger or smaller than the referent number in physical size. A NDE was obtained for the intentional task condition while a size congruity effect was obtained for the non-intentional task condition. The results of these studies suggest that numerical magnitude processing is independent of task, as the markers of magnitude processing are observed in both intentional and non-intentional task conditions.

Yet, other findings suggest that numerical magnitude processing is task-dependent. For example, Ganor-Stern and Tzelgov (2008) reported a NDE for Indian and Arabic numbers for the Numerical Matching Task, but not the Physical Matching Task. Similarly, Goldfarb et al.

(2011) reported a NDE for single digit Arabic numbers in the Magnitude Comparison Task, but not the Numerical Matching Task. Goldfarb et al. (2011) further proposed that whether participants process numerical magnitude depends on task demands, specifically, whether the task entails deep processing of quantity. Taken together, these results indicate that the extent to which task influences magnitude processing remains unclear.

1.2 The Influence of Notation on Numerical Magnitude Processing

Humans are generally able to recognise and process numbers belonging to a variety of notations and modalities. Examples of notations include symbolic (e.g., "5") and non-symbolic (e.g., "*****") numbers, whereas examples of modalities include auditory (e.g., "/faɪv/") and written numbers. As a case in point, written number words are diverse and include number words from alphabetic scripts (e.g., English, German, Malay), logographic scripts (e.g., Chinese, Japanese Kanji, Korean Hanja) and more. Thus, exactly which notations one is familiar with depends on one's cultural and linguistic background.

The extent to which notation influences numerical magnitude processing is under theoretical and empirical debate. Most theoretical models propose that magnitude representations are independent of notation. These include the Abstract Code Model (McCloskey, 1992), the Multiroute Model of Number Processing (Cipolotti & Butterworth, 1995), and the Triple Code Model (Dehaene & Cohen, 1995).

According to the Abstract Code Model (McCloskey, 1992; McCloskey & Macaruso, 1995), the mental systems that serve numerical cognition include the number comprehension systems for Arabic and verbal numbers, the number production systems for Arabic and verbal numbers, an abstract internal representation system, and a calculation system. The system that is responsible for semantic representation of number (i.e., magnitude) is thought to be independent of notation.

The Multiroute Model of Number Processing (Cipolotti & Butterworth, 1995) extended the Abstract Code Model by adding numerical input, output, and asemantic transcoding systems. Like the Abstract Code Model, the internal magnitude representation system is thought to be abstract, or notation-independent.

The Triple Code Model (Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2003) proposes that there are three representational codes for numerical information. The visual Arabic code represents digits. The auditory verbal code represents number words. The analog magnitude code stores semantic information about numerical magnitude and is said to be abstract, or notation-independent (c.f., Pesenti & Andres, 2009).

In contrast to the three models discussed above, the Encoding Complex Model (Campbell & Epp, 2004) posits that numerical processing is dependent on encoding and retrieval processes for each notation. This is a model of numerical cognition which posits that there are visual codes for numerals in Arabic notation and the individual’s first language, a magnitude code, and verbal codes for spoken number words in the individual’s first and second languages. Similarly, recent theoretical commentaries have highlighted various aspects of numerical notation that might affect processing, including the syntactic structure and orthographic properties of the notation or language (Daroczy, Wolska, Meures, & Nuerk, 2015; Dowker & Nuerk, 2016).

At an empirical level, there is debate regarding the type of task that is best suited to assess the effect of notation on numerical magnitude processing (e.g., Algom, 2009; Cohen Kadosh & Walsh, 2009). Furthermore, the results are inconsistent across tasks. Behavioural studies employ the logic that, to the extent that the processing of numerical magnitude is notation-independent, there will not be an interaction between notation and the NDE (Campbell & Epp, 2004; Cohen Kadosh & Walsh, 2009; Ito & Hatta, 2003). However, if processing of numerical magnitude is notation-dependent, there will be an interaction between notation and the NDE. This is based on the reasoning that, when two factors are independent of each other, manipulating one of them would result in an additive effect, or no overall effect; however, for two factors that depend on each other, manipulating one of them would result in an interaction (Cohen Kadosh & Walsh, 2009; Sternberg, 1969).

Applying this logic to examine tasks of magnitude processing yields mixed findings. With respect to intentional tasks of magnitude processing, some studies have reported no interaction between notation and the NDE, which supports a notation-independent view of magnitude processing. For example, Ito and Hatta (2003) employed the numerical size condition of the Numerical Stroop Task for Arabic, Kanji, and Kana numbers and did not find an interaction between notation and the NDE. Pinel et al. (2001) employed the Magnitude Comparison Task for Arabic and English notations and also did not find an interaction between notation and the NDE.

Others report an interaction between notation and markers of magnitude processing in intentional tasks, which is in favour of the notation-dependent account. Using a Magnitude Comparison Task for Arabic and Chinese numbers, Campbell et al. (1999) reported an interaction between notation and the NDE. Using a larger / smaller than 5 task, various studies reported an interaction between notation and the NDE: Cao et al. (2010) for Arabic and Chinese numbers, Cohen Kadosh (2008) for Arabic and Hebrew numbers, and Lukas et al. (2014) for Arabic, English, and Turkish numbers. Using the numerical size condition of the Numerical

Stroop Task for Arabic and Indian numbers, Ganor-Stern and Tzelgov (2008; Experiment 1) reported an interaction between notation and the NDE.

With respect to non-intentional tasks of magnitude processing, the results are mixed. For the Numerical Matching task, existing studies have not reported an interaction between notation and the NDE. This supports the notation-independent account of magnitude processing. Using Arabic numerals and English number words in this task, Dehaene and Akhavein (1995) did not find a significant interaction between notation and the NDE. Using Arabic and Indian numerals in this task, Ganor-Stern and Tzelgov (2008) also did not find an interaction between notation and the NDE. As for the Physical Matching task, Dehaene and Akhavein (1995) reported an interaction between notation and the NDE whereas Ganor-Stern and Tzelgov (2008) did not. However, in the physical size condition of the Numerical Stroop Task for Arabic and Hebrew numbers, Cohen Kadosh et al. (2008) reported an interaction between the size congruity effect and notation. Taken together, these findings suggest that the extent to which notation influences numerical magnitude processing remains unresolved.

1.3 The Current Study

The main aim of the study was to investigate the extent to which task and notation influence the processing of numerical magnitude, while extending previous work with respect to the research questions, research design, stimulus notations, and participant pool. Three research questions were developed:

1.3.1 To What Extent does Task Influence Numerical Magnitude Processing?

For the first research question, we investigated the extent to which numerical magnitude processing is dependent on or independent of task. In contrast to previous studies that investigated one or two tasks of numerical processing, the present study comprised three tasks:

Magnitude Comparison, Numerical Matching, and Language Matching Tasks. The Magnitude Comparison Task was regarded as an intentional task of magnitude processing as it requires participants to judge which number was larger in magnitude. The Numerical Matching Task was regarded as a non-intentional task because participants are not required to make a larger / smaller judgment; rather, they judge whether the numbers are numerically same or different. Finally, the Language Matching Task was deemed a non-intentional task as participants are not required to make a larger / smaller judgment; rather, they judge whether the language of the number words are the same or different.

Furthermore, in contrast to previous studies that adopted a between-subjects design (e.g., Ito & Hatta, 2003; Dehaene & Akhavein, 1995; Goldfarb et al., 2011), we selected a within-subjects design, where the same group of participants completed all three tasks. We reasoned that a within-subjects design would be more statistically powerful and would allow us to draw stronger conclusions from a cross-task comparison than a between-subjects design.

To the extent that the processing of numerical magnitude is dependent on task, we expected a two-way statistical interaction between Task and Numerical Distance. However, if numerical magnitude processing is independent of task, we expected that there would not be an interaction between Task and Numerical Distance. Based on the findings of related studies (Ganor-Stern & Tzelgov, 2008; Goldfarb et al., 2011; but see Dehaene & Akhavein, 1995), we hypothesized that numerical magnitude processing would be task-dependent.

1.3.2 To What Extent does Notation Influence Numerical Magnitude Processing?

For the second research question, we investigated the extent to which numerical magnitude processing is dependent on or independent of notation. Our study employed pure English (e.g., nine four), pure Chinese (e.g., 九 四), mixed Chinese-English (e.g., 九 four), and mixed English-Chinese (e.g., nine 四) notations. Although all the stimuli are written number words, the English and Chinese orthographies are different: English is an alphabetic script where several letters combine to form a unit of meaning in a word, while Chinese is a logographic script where each character represents a unit of meaning. Whilst Arabic notation has been extensively investigated on its own and in combination with other notations, the stimulus pairing of English and Chinese number words only have not yet been investigated in the existing literature.

In contrast to previous studies that presented stimuli in blocks of one notation condition at a time (e.g., Cohen et al., 2013; Cohen Kadosh et al., 2008, Experiment 1; Pinel et al., 2001), we randomized the notation of stimulus presentation such that trials from all notations were mixed within a single block. To date, few studies have presented more than one notation at a time in a single block (e.g., Cohen Kadosh et al., 2008, Experiment 3; Ganor-Stern & Tzelgov, 2008). Randomizing the notation of stimulus presentation has the advantage of reducing the likelihood of stimulus predictability, as participants will not know beforehand the notation of each trial.

Furthermore, in contrast to previous studies that did not account for possible differences in participants' fluency and experience with the stimulus notations, we incorporated a language background questionnaire to screen potential participants. Based on their responses, we selected individuals who were bilingual in English and Chinese with a balanced profile of language dominance for both languages to participate in the experiment. Such an approach would allow us to rule out the possibility that any notation differences found might have been due to intraindividual differences in participants' fluency and experience with the two notations.

In spite of obtaining a sample of balanced bilinguals with which to conduct the study, it could be argued that participants' language dominance across multiple contexts in daily life (e.g., reading, writing, speaking) might be different from language dominance for numerically-oriented settings. To address the possibility that the latter may have affected the experimental results, we presented participants with a Spontaneous Counting Task to ascertain their preferred language for numerical counting, and included this as an independent variable in the preliminary analyses.

To the extent that the processing of numerical magnitude is dependent on stimulus notation, we expected a two-way statistical interaction between Notation and Numerical Distance. However, if magnitude processing is independent of stimulus notation, we expected there would not be an interaction between Notation and Numerical Distance. As previously discussed, the findings of related studies are mixed—some report notation-dependence (e.g., Campbell et al., 1999; Cao et al., 2010; Lukas et al., 2014), others report notation-independence (e.g., Dehaene & Akhavein, 1995). Thus, we carefully re-examined this issue using multiple tasks in a within-subjects design. Based on the findings of studies that were most closely related to ours, we hypothesized that numerical magnitude processing would be notation-dependent (Campbell et al., 1999; Cao et al., 2010; Dehaene & Akhavein, 1995).

1.3.3 Do Task and Notation Interact to Influence Numerical Magnitude Processing?

For the third research question, we investigated whether the factors of task and notation have an interactive influence on magnitude processing or not. For example, the effect of notation on magnitude processing could be consistent across tasks, or it could be dependent on task. Conversely, the effect of task on magnitude processing could be independent of or dependent on notation.

To date, studies that have tested multiple task and notation conditions have not explicitly examined this issue. Rather, the analyses were split by task (e.g., Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008); thus, there are no conclusive findings on whether task and notation have an interactive influence on magnitude processing or not. For example, using Arabic, Indian, and mixed notation, Ganor-Stern and Tzelgov (2008) found an NDE for the Numerical Matching Task for all notations, but no NDE for the Physical Matching Task for all notations. Using Arabic, English, and mixed notation, Dehaene and Akhavein (1995) reported an interaction between notation and the NDE for the Physical Matching Task but not the Numerical Matching Task. Both patterns of results suggest that task and notation may have an interactive effect on magnitude processing, but this needs to be examined more directly.

In examining the possibility of an interactive influence of task and notation on magnitude processing, we considered participants' strategies in completing the task, and whether this might differ according to notation. For the Magnitude Comparison Task, we expected that participants would intentionally process numerical magnitude regardless of notation. This would be indicated by a NDE for all notations. For the Language Matching Task, participants do not necessarily have to process numerical magnitude and can use other non-numerical strategies, like matching the numbers based on their visual properties. Exactly which strategy participants employ might depend on the speed of processing of the strategies (Schwarz & Ischebeck, 2003). Nevertheless, we expected that participants would not use a strategy based on numerical magnitude because it was not required by the task.

It was unclear as to which strategy participants might employ for the Numerical Matching Task. One perspective is that participants might employ a strategy based on numerical magnitude for the mixed (i.e., different notation) trials, but not the pure (i.e., same notation) trials. For pure trials, participants can either complete this task using a strategy based on numerical magnitude (whether one number is larger than another), numerical value (same / different categorical judgment) or by visually matching the shape of the numbers (Wong & Szucs, 2013). For mixed trials, it is thought that participants are likely to match the numbers based on their quantity (van Opstal & Verguts, 2011). Another perspective, stemming from the bilingual language dominance literature, is that participants could use a non-numerical strategy to match the mixed trials. According to the Revised Hierarchical Model (Kroll & Stewart, 1994), it is thought that bilinguals have direct lexical links between their first and second languages, without necessarily accessing the meaning of the word. It seems possible that the pool of balanced bilinguals in this study could employ a lexical strategy for the mixed notation trials in the Numerical Matching Task by converting the number words into the same language, and then use a perceptual strategy to match the numbers based on their visual shape. We were therefore interested to investigate whether participants would employ a strategy based on numerical magnitude or not for the mixed trials in the Numerical Matching Task.

To the extent that task and notation have an interactive effect on magnitude processing, we expected a three-way statistical interaction between Task, Notation, and Numerical Distance. However, if task and notation do not interact to affect magnitude processing, we expected that there would not be a significant three-way interaction.

2. Method

2.1 Participants

152 right-handed participants aged 19 – 39 years who were bilingual in English and Chinese (103 females) participated in the online Language Background Questionnaire. Full questionnaire materials are publicly accessible at our project’s webpage on the Open Science

Framework (https://osf.io/hk3wz/). Participants were recruited via word of mouth, online newsletters to university students and staff, and posters around the campus. All provided informed consent online and indicated their interest to participate in the rest of the study. Based on responses to the questionnaire, participants were classified into three language dominance groups: English-dominant, Chinese-dominant, and balanced bilinguals. Full coding and scoring procedures are available on our project's webpage on the Open Science Framework (https://osf.io/hk3wz/). Those who were deemed to show a balanced profile of language dominance in English and Chinese were invited to participate in the rest of the experiment.

Forty participants completed the full study. Outlier participants were defined as those individuals whose average reaction times (RT) differed more than +/- 3 standard deviations (SD) from the average RT of all participants. There was 1 outlier participant; therefore a total of 39 participants were included in the analyses (25 females, *M* = 26.44 years, *SD* = 4.63). All provided written informed consent. The study was approved by the Institutional Review Board of the University affiliated to the first and second authors.

2.2 Procedure

Each participant completed all tasks during two individual testing sessions. The sessions were held at a quiet setting within the University. All tasks were computerized and presented on a 15 inch laptop computer screen, with participants sitting approximately 50cm away from the screen. E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA) was used to programme and present the tasks.

2.3 Measures

2.3.1 Language Background Questionnaire

The aim of the questionnaire was to ascertain respondents' language dominance profiles. The measures, described below, were adapted from previously validated language dominance questionnaires, namely, Dunn and Tree (2009), Gertken, Amengual, and Birdsong (2014), and Lim, Liow, Lincoln, Chan, and Onslow (2008).

2.3.1.1 Language History.

Participants stated the age of acquisition for English and Chinese, for both formal and informal learning, as well as the language in which they currently use for basic mathematics.

2.3.1.2 Subjective Language Proficiency (English and Chinese).

Participants rated their proficiency in understanding, speaking, reading, and writing in English and Chinese on a scale of 0 (very few words) to 6 (native proficiency). Internal consistency was acceptable for both English and Chinese (Cronbach's $\alpha \geq .94$).

2.3.1.3 Frequency of Use (English and Chinese).

Participants indicated how frequently they typically use English relative to Chinese in various modalities such as speaking, reading, and writing in daily life, and in various contexts such as at home, at work / school, and in social settings. Internal consistency was acceptable for both English and Chinese (Cronbach's $\alpha \geq .84$)

2.3.1.4 Objective Language Proficiency.

Participants indicated their grades for English and Chinese for the national examinations at their final year of Secondary schools (16 to 17 years) and Junior College (18 years). Participants who did not study in local schools had the option of providing their academic grades for their national equivalent.

2.3.1.5 Language Background for Early Mathematics Learning

To assess participants' background regarding the language with which they first learnt mathematics formally and informally, participants stated their nationalities, the language in which they learnt mathematics in elementary school, as well as the language in which they first learnt mathematics at home.

2.3.2 Spontaneous Counting Task.

Participants were presented with a display of non-overlapping items. They were instructed to count aloud, in their preferred language, the items one by one, and then to say aloud the total number of items. Participants were instructed to respond as quickly and as accurately as possible. The instructions for this task were administered in both English and Chinese. The experimenter noted down their accuracy and the language in which they counted.

2.3.3 Numerical Processing Tasks.

A fully within-subjects Task (3) x Notation (4) x Numerical Distance (2) design was employed. The levels of task were: Magnitude Comparison, Numerical Matching, and Language Matching. The number words ranged from 2 to 10 and were presented in four levels of Notation: pure English, pure Chinese, mixed Chinese-English, and mixed English-Chinese. The levels of Numerical Distance were: Small (Distance = 1) and Large (Distance = 6, 7, or 8). Trials from all notations and distances were randomly mixed in each block.

2.3.3.1. Task Instructions and Design.

In all tasks, participants viewed two number words, one on the left side of the screen and the other on the right. Participants were told to place their left and right index fingers on two yellow stickers on a QWERTY keyboard to indicate a left or right button-press response respectively. In the Magnitude Comparison task, participants were instructed to make a button-press response corresponding to the spatial position of the number word that was larger in numerical magnitude. That is, if the number word on the left side of the screen was larger in magnitude, participants would make a left button-press response, and vice versa. In the Numerical Matching Task, participants were instructed to make a button-press response to indicate whether the numerical values of the two words were the same or different. In the Language Matching Task, participants were instructed to make a button-press response to indicate whether the languages of the two words were the same or different. In all tasks, participants were told to respond as quickly and as accurately as possible. For the Numerical and Language Matching Tasks, the left / right button assignment used to indicate same / different were counterbalanced within subjects, while the order of the button assignment and the order of the task were counterbalanced between subjects in a balanced Latin Square design.

Each task began with a practice block of 6 trials where participants were presented with feedback of their accuracy and RT at the end of each trial. This was followed by the experimental blocks, where participants were not given any feedback. For the Magnitude Comparison Task, there was one experimental block of 192 trials. For the Language and Numerical Matching Tasks, each task comprised two experimental blocks of 144 trials, giving rise to a total of 288 trials per task. 50% of trials in each block required a “same” response and the rest required a “different” response. The “different” number word pairs were identical for all tasks and the unique number word pairs are presented in Appendix A. The experimental schematic was identical across all tasks and is presented in Figure 1.

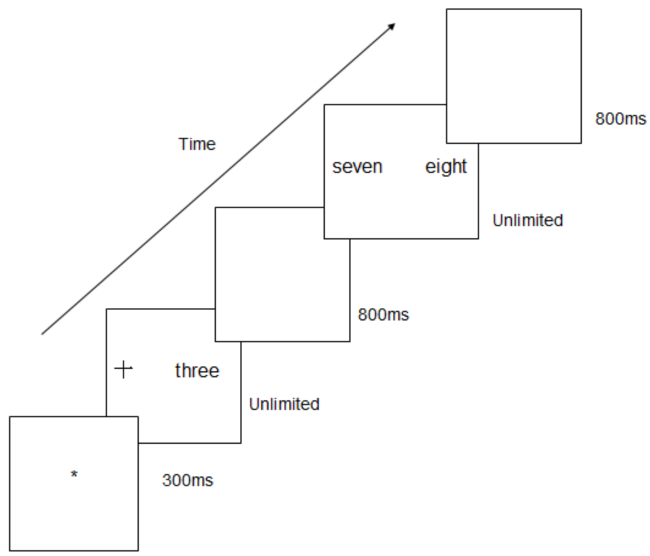

Figure 1. Schematic of numerical processing tasks. The example trials displayed are 10 3 in mixed Chinese-English notation and 7 8 in pure English notation.

2.3.3.2. Stimulus Notation.

Employing English and Chinese notations enabled us to investigate numerical processing for the quantity "10", which is represented by a single word / character in English ("ten") and Chinese ("十"), but consists of two digits in Arabic notation and is seldom investigated with other single-digit Arabic numbers (see Moeller, Pixner, Kaufmann, & Nuerk, 2009).

As seen in the Chinese number word pairs in Appendix A, it can be argued that the number words for 2 and 3 contain visual information about numerical size as they are represented by two and three strokes respectively. Nevertheless, these numbers were included in the stimulus set because the Chinese number words for 8, 9, and 10 are also represented by two strokes.

3. Results

3.1 Participants' Language Background for Early Mathematics Learning

Participants' nationalities included Singapore (*N* = 23), China (*N* = 8), Malaysia (*N* = 5), and Hong Kong (*N* = 3). Most participants had formal early mathematics experiences in the English language (*N* = 23) and the rest in Chinese (*N* = 16). Conversely, most participants had informal early mathematics experiences in the Chinese language (*N* = 28) and the rest in English (*N* = 11). Thus, participants had a heterogeneous background regarding the language with which they learnt mathematics formally and informally. As a result, the Spontaneous Counting Task was used to address participants' heterogeneous background regarding the language with which they learnt mathematics formally and informally.

3.1 Descriptive Statistics

Overall accuracy for the three tasks was close to ceiling (*M* = 96.14, *SD* = 2.53), as was accuracy for each task: Magnitude Comparison (*M* = 95.91, *SD* = 2.87), Numerical Matching (*M* = 96.60, *SD* = 2.16), and Language Matching (*M* = 95.90, *SD* = 2.53). Bonferroni-corrected bivariate correlations revealed that there were no significant speed-accuracy tradeoffs. Trials which differed more than +/- 3 SD from each participant's mean RT to correct trials, were deemed outliers. Outlier trials (1.67%) and incorrect trials (3.86%) were excluded from the data analyses. Given the ceiling effect for the accuracy data, we decided not to analyse the accuracy data. Instead, we analysed participants' adjusted RT scores by dividing mean RT by accuracy (Adjusted RT = RT / Acc). This is sometimes termed inverse efficiency and can be interpreted as a measure for RT after adjusting for inaccurate performance, thereby estimating how slow participants would respond if they responded with 100% accuracy (Astle, Jackson, & Swainson, 2012; Chan, Aw & Tang, 2011; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Khng & Lee, 2014). Given the doubts that have been raised about this measure (Bruyer & Brysbaert, 2011), we performed the same analyses below based on RT to correct trials. The trend of the data for both sets of results were very similar, with the adjusted RT data being slightly more conservative.

3.2 Preliminary Analyses and Overview

The purpose of the preliminary analysis was to ascertain whether (1) participants' preferred language for numerical counting, and / or (2) the order of mixed notation number words (i.e., Chinese-English vs English-Chinese) influenced the results. A 3 X 4 X 2 X 2 mixed ANOVA was performed on participants' adjusted RTs. The within-subjects variables were Task (Magnitude Comparison, Numerical Matching, Language Matching), Notation (Pure English, Pure Chinese, Chinese-English, and English-Chinese), and Numerical Distance (Small, Large). The between-subjects variable was Spontaneous Counting Language: English ($N = 21$) and Chinese ($N = 18$). The main effect of Spontaneous Counting Language was not statistically significant ($p = .27$). Furthermore, the results yielded no significant differences between Chinese-English and English-Chinese trials ($p = .79$). In light of the above, Spontaneous Counting Language was removed as a variable, and the English-Chinese and Chinese-English trials were collapsed to form a single level of notation, henceforth termed Mixed Notation.

For the analyses, we adopted an inclusive statistical approach by performing frequentist statistics on a 3 X 3 X 2 repeated-measures ANOVA on participants' adjusted RTs. Bayesian statistics were run in parallel on the same analysis. While frequentist statistics provide information about the long run probability of the data under the null hypothesis, one key advantage of Bayesian statistics is that they allow us to quantify evidence for a given hypothesis by providing us information about the probability of the hypothesis being true. In this study, Bayesian statistics afforded us the opportunity to quantify the support for the results that emerged from our frequentist analyses by way of a model comparison approach (e.g., Wagenmakers et al., 2017). Specifically, to assess the extent to which the data support the presence vs absence of any given term, it is possible to specify different statistical models and compare their relative predictive performance using the Bayes factor (Marsman & Wagenmakers, 2017).

The within-subjects variables in our study were: Task (Magnitude Comparison, Numerical Matching, Language Matching), Notation (Pure English, Pure Chinese, Mixed), and Numerical Distance (Small, Large). Mean adjusted RTs for all task and notation conditions are displayed in Figure 2.

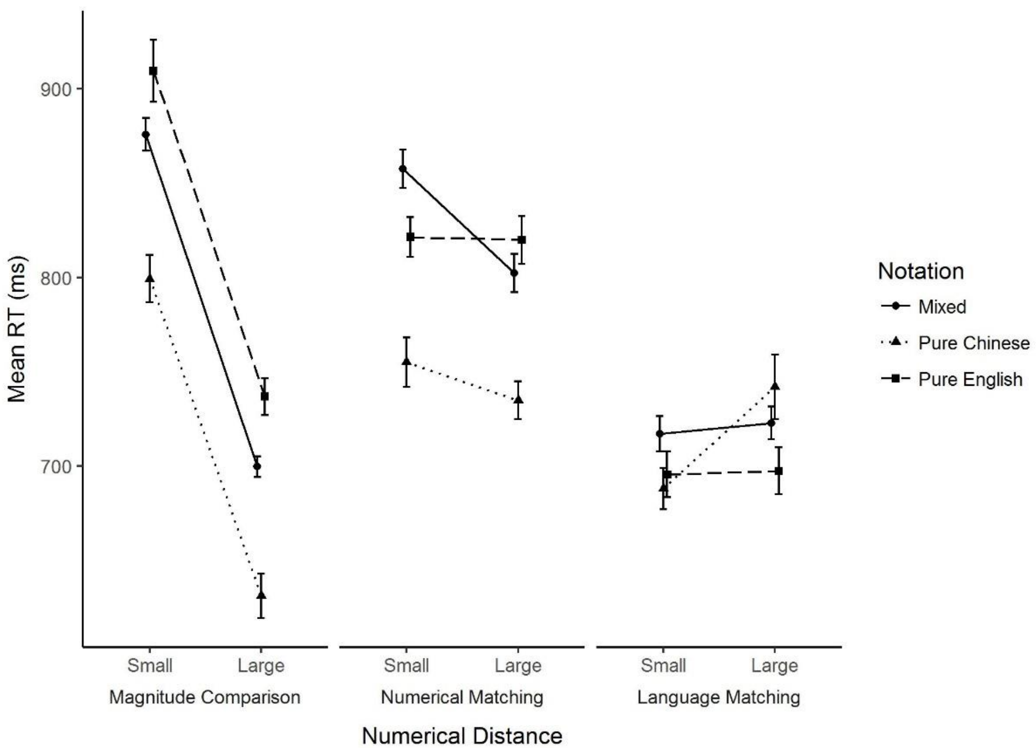

Figure 2. Mean adjusted RTs for all task and notation conditions. Error bars represent standard errors of the means.

3.3 Frequentist Statistics

All the main effects and interactions were significant. There was a significant main effect of Numerical Distance, F (1, 38) = 107.03, $p < .001$, $\omega^2 = .73$, where adjusted RT to large distance trials (731.94ms) was significantly faster than small distance trials (791.10ms). Next, there was a significant main effect of Notation, F (1.65, 62.71) = 23.80[2], $p < .001$, $\omega^2 = .37$, where adjusted RT was faster for pure Chinese trials (725.13ms) than pure English (780.17ms), and mixed notation trials (779.27ms). Follow-up pairwise comparisons indicated that adjusted RT to pure Chinese trials were significantly faster than pure English and mixed notation trials (all *p's* < .001), but the latter two notations did not differ significantly. There was also a significant main effect of Task, F (1.72, 65.41) = 14.74, $p < .001$, $\omega^2 = .26$, where adjusted RT was fastest for Language Matching (710.56ms), followed by Magnitude Comparison (775.44ms), and Numerical Matching (798.57ms). Follow-up pairwise comparisons indicated that adjusted RT to the Language Matching task was significantly faster than the Magnitude Comparison (p = .003) and Numerical Matching tasks ($p < .001$). However, adjusted RT to the latter two tasks did not differ significantly.

There was a significant interaction between Task and Numerical Distance, F (2, 76) = 110.70, $p < .001$, $\omega^2 = .74$, where the NDE was largest for the Magnitude Comparison task (172.30ms), followed by the Numerical Matching task (25.71ms). The distance effect was reversed for the Language Matching task (20.52ms). Simple effects analyses revealed that the NDE was significant for the Magnitude Comparison ($p < .001$) and Numerical Matching tasks (p = .002). The reversed distance effect was also significant for the Language Matching Task (p = .02). Next, there was a significant interaction between Task and Notation, F (2.40, 91.07) = 16.26, $p < .001$, $\omega^2 = .28$. For the Magnitude Comparison and Numerical Matching tasks, adjusted RT to pure Chinese trials were faster than pure English trials ($MC_{Eng - Chi}$ = 100.97ms,

[2] For analyses where the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied.

$NM_{Eng - Chi}$ = 75.63ms), and mixed trials ($MC_{Mixed - Chi}$ = 72.50ms, $NM_{Mixed - Chi}$ = 84.97ms). Simple effects analyses indicated that the RT advantage for pure Chinese trials was significant in both tasks (all *p*'s < .001). However for the Language Matching task, there was no significant effect of the notation. There was also a significant interaction between Notation and Numerical Distance, F (2, 76) = 4.33, p = .02, ω^2 = .08, where the NDE was largest for mixed notation (75.27ms), followed by pure English (57.44ms) and pure Chinese (44.78ms) notation. Simple effects analyses indicated that the NDE was significant for all notations (all *p*'s < .001).

Finally, there was a significant three-way interaction between Task, Notation, and Numerical Distance, F (4, 152) = 2.72, p = .03, ω^2 = .04. To break down the interaction, we tested the Notation x Numerical Distance interaction for each task. For the Magnitude Comparison task, the two-way interaction was not significant (p = .90). To test if the NDE was significant for each notation, we conducted pairwise comparisons followed by a Bonferroni correction. The results revealed that the NDE was indeed significant for all notations in this task (172.66ms for pure English, 168.10ms for pure Chinese, 176.14ms for mixed notation; all *p*'s < .001). For the Numerical Matching task, the two-way interaction was significant, F (1.67, 63.49) = 4.18, p = .03. Pairwise comparisons indicated that the NDE was significant for the mixed notation only (55.40ms; p < .001). For the Language Matching task, the two-way interaction was significant, F (1.54, 58.43) = 5.21, p = .01. Pairwise comparisons indicated that the reverse distance effect was significant for the Chinese notation only (55.40ms; p = .004).

3.4 Bayesian Statistics

To address our second and third research question, two interaction terms were of interest: Notation x Numerical Distance and Task x Notation x Numerical Distance respectively. In the frequentist statistics presented above, both interactions were significant but yielded extremely small effect sizes. Thus, to quantify support that the data offered for these interactions, we conducted a Bayesian repeated-measures ANOVA using JASP (Version 0.8.1.2; JASP Team, 2017). We employed a uniform prior distribution ranging from -1 to +1, in which every value of p in this range is equally plausible a priori (Marsman & Wagenmakers, 2017). We employed the procedures of Wagenmakers et al. (2017; also see Marsman & Wagenmakers, 2017) to compare model fit under the null model of three main effects and alternative candidate models containing three main effects plus various combinations of two- and three-way interactions, elaborated below.

First, we specified the null and candidate models to examine their Bayes Factors (BF). The null model contained three main effects. To quantify support for the three-way interaction (Task x Notation x Numerical Distance), we examined Model 1, which contained all possible main effects and two- and three-way interactions. To quantify support for the interaction between Notation x Numerical Distance, we examined Model 2, which contained all possible main effects and two-way interactions. Model 3 contained all possible main effects and two-way interactions that excluded Notation x Numerical Distance and Task x Notation x Numerical Distance – the terms that yielded very small effect sizes. Models 4 – 8 were alternative models containing other combinations of the two-way interactions. To interpret the size of the Bayes factors for each model, we used the recommendation of Jeffreys (1961), which proposes that a BF of over 100 constitutes decisive evidence for the model. Full details of all models are available in Appendix B. Model 3 best predicted the data (BF = $1.68*10^{60}$).

Next, using the principle of transitivity, we arbitrated between several models by dividing the posterior probability of Model 3 by the posterior probability of each alternative model (Jeffreys, 1961; Marsman & Wagenmakers, 2016). The purpose of this analysis was to assess the strength of evidence in favour of Model 3 over the null model, and models containing the key interaction terms (i.e., Models 1 & 2). To compare Model 3 and the null model, we divided the posterior probability of Model 3 by that of the null model (0.88 / $3.34*10^{-30}$). This revealed decisive evidence that Model 3 was preferred to the null model (BF = 2.63 $*10^{29}$). To assess the importance of the three-way interaction, we compared Model 3 against Model 1 (0.88 / 0.009). This yielded decisive evidence against the inclusion of the three-way interaction (BF = 97.78). To assess the importance of the interaction between Notation x Numerical Distance, we compared Model 3 against Model 2 (0.88 / 0.12). This yielded moderate evidence against the inclusion of Notation x Distance (BF = 7.33).

4. Discussion

The extent to which task and notation influence the processing of numerical magnitude is under theoretical and empirical debate. To date, behavioural studies have yielded a mixed body of evidence. Here, we re-examined this issue whilst extending the literature by: (1) investigating whether task and notation have interactive effects on numerical magnitude processing or not; (2) administering multiple tasks of numerical processing in a within-subjects design; (3) using a unique combination of stimuli that were not investigated together in past research (pure English, pure Chinese, and mixed notation number words); (4) accounting for potential intra-individual differences in fluency and experience with the stimulus notations by recruiting participants who were bilingual in English and Chinese with a balanced profile of language dominance for both languages; and (5) randomizing the notation stimulus presentation, thus eliminating stimulus predictability as a possible methodological concern. We employed the NDE as an index of numerical magnitude processing.

4.1 Numerical Magnitude Processing is Task-Dependent

In response to the first research question, the results indicated that the numerical magnitude processing of written number words is dependent on task. In the frequentist statistics, this was evidenced by a significant two-way interaction between Task and Numerical Distance. Specifically, the NDE was significant in the Magnitude Comparison and Numerical Matching tasks, but this effect was reversed for the Language Matching Task. The importance of the interaction between Task and Numerical Distance was confirmed by the Bayesian statistics, which revealed that the model that best predicted our data had three main effects (Task, Notation, Numerical Distance) and two interactions (Task x Numerical Distance, Task x Notation). We discuss the results for each task below.

The significant NDE in the Magnitude Comparison Task indicates that participants processed numerical magnitude in this task. This concurs with previously reviewed studies that report the NDE in intentional tasks of magnitude processing for number words in various languages (e.g., Campbell et al., 1999; Cao, Li, & Li, 2010; Campbell et al., 1999; Cohen Kadosh, 2008; Cohen et al., 2013; Lukas et al., 2014; Pinel et al., 2001; Szucs & Csepe, 2005). Thus, our results confirm and offer behavioural evidence for intentional magnitude processing for pure English, pure Chinese, and mixed notation number words using the Magnitude Comparison task.

The significant NDE in the Numerical Matching Task indicates that, overall, participants processed numerical magnitude in this task. Subsequent analyses found a different pattern of results depending on notation, and indicate that the NDE in this task was driven by mixed but not pure notation trials. This pattern of results concurs with previously reviewed studies that report the NDE among mixed notation trials in the Numerical Matching Task (Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008; Van Opstal & Verguts, 2011; Verguts & Van Opstal, 2005) but not pure notation trials (Goldfarb et al., 2011; Wong & Szucs, 2013).

There was no significant NDE in the Language Matching Task, which suggests that participants did not process numerical magnitude in this task. Instead, an overall reverse distance effect was found. The interpretation of this will be taken up in a later section. With regard to related studies that employed the Physical Matching Task, Dehaene and Akhavein (1995) reported a significant NDE for pure Arabic and pure English numbers, whereas our study and that of Ganor-Stern and Tzelgov (2008) did not. One possibility for the disparity in findings could be the smaller sample size of 10 participants in Dehaene and Akhavein (1995), as compared with the larger sample size of 32 participants in Ganor-Stern and Tzelgov (2008) and 39 participants in our study.

In sum, our results indicate that numerical magnitude processing is task-dependent. Magnitude processing consistently occurred in the Magnitude Comparison task, and in mixed notation trials in the Numerical Matching Task. These results are similar but not identical to those of Goldfarb et al. (2011) and Ganor Stern and Tzelgov (2008), who found a NDE for some task conditions but not others. On this basis, we suggest that future research investigating the processing of numerical magnitude carefully consider the task and context in which this occurs.

4.2 Numerical Magnitude Processing is Notation-Independent

In response to the second research question, the results supported a notation-independent account of magnitude processing. The frequentist approach seemed to lend support to the notation-dependent account, based on a significant two-way interaction between Notation and Numerical Distance. However, the effect size for this interaction was extremely small ($\omega^2 = .08$). We further inspected the results for the Magnitude Comparison Task, which was the only task that consistently elicited numerical processing, as evidenced by a NDE for all notations. Notation and Numerical Distance did not interact in this task, indicating that the processing of numerical magnitude is notation-independent. Furthermore, upon re-examining the dataset with a Bayesian repeated-measures ANOVA, the analyses yielded moderate evidence against the inclusion of the two-way interaction between Notation and Numerical Distance. In other words, the effect of including a model with the Notation*Numerical Distance term on the posterior distribution was minimal. Thus, contrary to our hypothesis, our results favour a notation-independent account of magnitude processing.

Our results echo other empirical studies that did not find an interaction between Notation and the NDE for intentional tasks of numerical processing (Ito and Hatta, 2003; Pinel et al., 2001) and the Numerical Matching Task (Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008). Also, our results are in support of theoretical models that advocate a notation-independent account of magnitude processing, such as the Abstract Code Model (McCloskey, 1992), Multiroute Model of Number Processing (Cipolotti & Butterworth, 1995), and the Triple Code Model (Dehaene & Cohen, 1995).

It is not possible to definitively resolve the discrepancy between our findings and those of others that report an interaction between Notation and Numerical Distance (e.g., Campbell et al., 1999; Cao et al., 2010; Cohen Kadosh, 2008; Cohen Kadosh et al., 2008; Lukas et al., 2014). However, we propose three factors that could account for this. First, existing studies tend to investigate the effect of notation using one or two tasks of magnitude processing. This has led to some debate over whether the nature of magnitude processing is best investigated under intentional or non-intentional task conditions. To circumvent this debate, we examined the effect of notation under three tasks of numerical processing. Our conclusions are based on examining the results for the tasks separately (particularly the Magnitude Comparison Task) and together.

Second, studies that report a significant interaction between Notation and Numerical Distance seldom report effect sizes. It is possible that their effect sizes were too small to be meaningful. The present study examined the effect sizes for the significant two-way interaction involving Notation and Numerical Distance, and also examined the data using a Bayesian approach. This allowed us to reject a notation-dependent account of magnitude processing based on the data in our study.

Third, methodological concerns may have affected the experimental results of some existing studies. We previously noted that most experiments do not account for possible differences in participants' fluency and experience with the stimulus notations, which may have introduced a processing bias in favour of the more frequently used notation. For example, it is possible that a smaller NDE to the more fluent or frequently used notation may occur, because participants have more precise magnitude representations with less distributional overlap for that notation, leading to the finding of an interaction between Notation and Numerical Distance. The present study addressed this concern by recruiting individuals who were bilingual in English and Chinese with a balanced profile of language dominance for both languages, and tested them on English, Chinese, and mixed notation number words. Also, the results from the Spontaneous Counting Task confirmed that participants' preferred language for numerical counting did not have an impact on these findings. Furthermore, by randomizing the notation of stimulus presentation, our study eliminated the potential methodological concern of stimulus predictability. Addressing these issues strengthened our conclusions of a notation-independent account of magnitude processing, based on the data in our study.

4.3 Task and Notation Do Not Interact to Affect Magnitude Processing

In response to the third research question, the results indicated that task and notation do not interact to affect magnitude processing of written number words. This was evidenced by the lack of strong evidence for a three-way interaction between Task, Notation, and Numerical Distance. Although the frequentist approach yielded a significant three-way interaction, the p-value ($p = .03$) and effect size ($\omega^2 = .04$) did not constitute strong evidence in support of this interaction. Upon breaking down this interaction, we found that magnitude processing occurred for all notation conditions in the Magnitude Comparison Task, and the mixed notation condition for the Numerical Matching Task only. Additionally, upon examining the data with a Bayesian repeated-measures ANOVA, the analyses yielded decisive evidence against the inclusion of the three-way interaction. On this basis, we propose that task and notation do not interact to affect magnitude processing.

It is likely that the follow-up results of the interaction point to differences in participants' strategies in completing each task. With regard to the Magnitude Comparison Task, participants processed numerical magnitude regardless of notation, as indicated by a significant NDE for all notations. For the Numerical Matching Task, participants processed numerical magnitude for the mixed notation trials, but not the pure notations. This was indicated by a significant NDE for the mixed notation trials only, and is in line with previous findings that participants can use strategies that are not based on numerical magnitude to handle pure notation trials (Wong & Szucs, 2013), but handle mixed notation trials based on numerical magnitude (van Opstal & Verguts, 2011). The finding that mixed and pure notation trials were processed differently in this task is noteworthy, given that the notation for stimulus presentation was randomized. In other words, although participants employed a magnitude processing strategy for the mixed notation trials, this strategy did not "spill over" into the pure notation trials, even though both trial types were inter-mixed and randomly presented for this task. For the Language Matching Task, it is likely that participants did not use a strategy based on numerical magnitude, as evidenced by the lack of a significant (forward) NDE for the notations.

4.4 Reverse Distance Effect: Artifact or True Effect?

The reverse distance effect that occurred for Chinese trials in the Language Matching Task is difficult to interpret and was not predicted. It is possible that this result is simply an artifact, because it emerged in the adjusted RT data, but not the non-adjusted RT data. In other words, this observed effect may have been mainly driven by the incorrect trials for this condition. Furthermore, reverse distance effects typically occur in the context of numerical order processing (Turconi, Campbell, & Seron, 2006) or among individuals with mathematical learning difficulties (Rouselle & Noel, 2007; also see Bachot, Geverse, Fias, & Roeyers, 2005)—neither of which were relevant in the current task.

Having said this, several studies have similarly reported reverse distance effects in non-intentional tasks of numerical magnitude (e.g., Girelli et al., 2000; Tang, Critchley, Glaser, Dolan, & Butterworth, 2009; Tzelgov et al., 1992). However, for these studies, the reverse distance effect occurred in the context of a Numerical Stroop Task and so the interpretation of their results might not be directly applicable to our study.

4.5 Task and Notation Interact to Affect Speed of Processing

Finally, the results also indicate that task and notation have an interactive effect on speed of processing. In the frequentist statistics, this was evidenced by a significant two-way interaction between Notation and Task. Follow up analyses revealed that participants responded to pure Chinese trials more quickly than other notations for the Magnitude Comparison and

Numerical Matching tasks. However, the RT advantage to pure Chinese trials was lost for the Language Matching task. In the Bayesian statistics, the model that best predicted the data confirmed the important role of the two-way interaction between Notation and Task.

These findings suggest that linguistic factors, combined with task, may have an overall impact on speed of processing. For example, it is possible that the Chinese orthography has a greater economy in conveying meaning than the English orthography and the mixed notation. Chinese is a logographic script in which each symbol (a Chinese character) concisely conveys a unit of meaning. In contrast, English is an alphabetic script where a group of symbols (several English letters) combine to convey a unit of meaning. For example, the quantity "7" is conveyed by one Chinese character (七) but five English letters (seven). Similarly, the phonology of the Chinese number words has a greater economy than that for English number words. This is because the Chinese numbers from 2 to 10 contain, on average, a smaller number of phonemes and syllables than the equivalent English numbers.

It is possible that, when participants have to process at some level the meaning of the stimuli (e.g., large / small in the Magnitude Comparison Task; same / different in the Numerical Matching Task), the economy of the orthography and / or phonology makes a difference to participants' speed of processing. However, when participants are not required to process the meaning of the stimuli, and can use a visual / verbal strategy based on the physical properties of the stimuli (as in the Language Matching Task), it is possible that differences in the economy of the orthographies / phonologies no longer makes a difference to participants' speed of processing. This is one possible explanation for our results, and further research may examine this phenomenon further, including testing other notations such as number words from other languages, or even romanised Chinese (Hanyu Pinyin) and a larger stimulus set, such as multidigit numbers. Nevertheless, based on our data, one implication for future research investigating the influence of numerical notation is to consider the task conditions and context in which the study occurs.

4.6 Conclusions

In this study, we reported behavioural evidence that magnitude processing for written number words is influenced by task, rather than notation conditions. Specifically, magnitude processing was found to be task-dependent and occurred for the Magnitude Comparison Task and for mixed notation trials in the Numerical Matching Task only. Next, task and notation had an interactive influence on overall speed of processing, where participants responded to Chinese number words significantly faster than other notations for the Magnitude Comparison and Numerical Matching Tasks only. This suggests that linguistic factors, when combined with task, may interact to affect overall speed of processing (c.f. Dowker & Nuerk, 2016).

The Bayesian analyses afforded us the opportunity to examine two terms that yielded small effect sizes in the frequentist statistics (i.e., Task x Notation x Numerical Distance, and Notation x Numerical Distance. The analyses yielded evidence against the inclusion of both terms. In other words, task and notation did not interact to affect magnitude processing; also, in response to the ongoing debate on the extent to which magnitude processing is abstract or not, the results support a notation-independent account of numerical magnitude processing.

Theoretical frameworks and commentaries in the field of numerical cognition have traditionally focused on the role of notation and its impact on numerical magnitude (see Cohen Kadosh & Walsh, 2009; Dowker & Nuerk, 2016). Our results, along with those of several others (e.g., Goldfarb et al., 2011) suggest that the role of task may be a more critical factor to consider in its impact on magnitude processing. Nevertheless, our results, when taken in the context of other empirical studies, represent but one perspective based on one type of research methodology with a specific set of number pairs and notation pairing. Continued behavioural and neuroimaging research into these research questions will enrich our understanding of the cognitive and neural mechanisms that underlie the nature of numerical magnitude processing across human development.

Appendix A

Unique Number Word Pairs Requiring a "Different" Response

Pure Chinese Notation	Pure English Notation	Mixed Chinese - English Notation	Mixed English - Chinese Notation	Numerical Distance
四 五	four five	四 five	four 五	1
五 六	five six	五 six	five 六	1
六 七	six seven	六 seven	six 七	1
七 八	seven eight	七 eight	seven 八	1
八 九	eight nine	八 nine	eight 九	1
九 十	nine ten	九 ten	nine 十	1
二 十	two ten	二 ten	two 十	8
二 九	two nine	二 nine	two 九	7
三 十	three ten	三 ten	three 十	7
四 十	four ten	四 ten	four 十	7
三 九	three nine	三 nine	three 九	6
二 八	two eight	二 eight	two 八	6

Note: This table presents the unique number pairs only. In the final stimulus list, the left vs right assignment of each number pair was randomized such that the larger number appeared on the left side on 50% of trials (Magnitude Comparison Task) / "different" trials (Numerical & Language Matching Tasks) and appeared on the right side for the other 50% of trials / "different" trials respectively.

Appendix B

Bayes Factors for Model Comparison

Model		Bayes Factor	Posterior Probability
Null Model (3 main effects)	Task + Notation + Numerical Distance	$6.39*10^{30}$	$3.34*10^{-30}$
Model 1 (3 main effects, 3 two-way interactions, 1 three-way interaction)	Task + Notation + Numerical Distance + Task x Notation + Task x Numerical Distance + Notation x Numerical Distance + Task x Notation x Numerical Distance	$1.77*10^{58}$	0.009
Model 2 (3 main effects, 3 two-way interactions)	Task + Notation + Numerical Distance + Task x Notation + Task x Numerical Distance + Notation x Numerical Distance	$2.21*10^{59}$	0.12
Model 3 (3 main effects, 2 two-way interactions)	Task + Notation + Numerical Distance + Task x Notation + Task x Numerical Distance	$1.68*10^{60}$	0.88

Model 4 (3 main effects, 2 two-way interactions)	Task + Notation + Numerical Distance + Task x Notation + Notation x Numerical Distance	$1.74*10^{34}$	$9.09*10^{-27}$
Model 5 (3 main effects, 2 two-way interactions)	Task + Notation + Numerical Distance + Task x Numerical Distance + Notation x Numerical Distance	$3.32*10^{53}$	$1.73*10^{-7}$
Model 6 (3 main effects, 1 two-way interaction)	Task + Notation + Numerical Distance + Notation x Numerical Distance	$6.20*10^{29}$	$3.24*10^{-31}$
Model 7 (3 main effects, 1 two-way interaction)	Task + Notation + Numerical Distance + Task x Numerical Distance	$2.38*10^{54}$	$1.25*10^{-6}$
Model 8 (3 main effects, 1 two-way interaction)	Task + Notation + Numerical Distance + Task x Notation	$1.60*10^{35}$	$8.36*10^{-26}$

References

Algom, D. (2009). Slippery platform: The role of automatic and intentional processes in testing the effect of notation. Commentary on Cohen Kadosh, R. & Walsh V., Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*, 328-329.

Astle, D. E., Jackson, G. M., & Swainson, R. (2012). Two measures of task-specific inhibition. *Quarterly Journal of Experimental Psychology*, *65*(2), 233-251.

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology Science*, *47*(1), 172.

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5-13.

Campbell, J. I., & Epp, L. J. (2004). An encoding-complex approach to numerical cognition in Chinese-English bilinguals. *Canadian Journal of Experimental Psychology, 58*(4), 229.

Campbell, J. I., Kanz, C. L., & Xue, Q. (1999). Number processing in Chinese-English bilinguals. *Mathematical Cognition*, *5*(1), 1-39.

Cao, B., Li, F., & Li, H. (2010). Notation-dependent processing of numerical magnitude: electrophysiological evidence from Chinese numerals. *Biological Psychology, 83*, 47-55.

Chan, W. W. L., Au, T. K., & Tang, J. (2011). Exploring the developmental changes in automatic two-digit number processing. *Journal of Experimental Child Psychology*, *109*(3), 263-274.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, *124*, 375–390.

Cohen, D. J., Warren, E., & Blanc-Goldhammer, D. (2013). Cross-format physical similarity effects and their implications for the numerical cognition architecture. *Cognitive Psychology*, *66*, 355-379.

Cohen Kadosh, R. (2008). Numerical representation: Abstract or nonabstract? Q*uarterly Journal of Experimental Psychology*, *61*(8), 1160-1168.

Cohen Kadosh, R, Brodsky, W., Levin, M., & Henik, A. (2008). Mental representation: What can pitch tell us about the distance effect? C*ortex*, *44*(4), 470-477.

Cohen Kadosh, R., Henik, A., & Rubinsten, O. (2008). Are Arabic and verbal numbers processed in different ways? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1377.

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences, 32*, 313 – 373.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *6,* 348.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*, 145 – 147.

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

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. In Butterworth, B. (Ed.), *Mathematical Cognition, Volume 1* (pp. 83–120). UK: Psychology Press.

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

Dowker, A., & Nuerk, H. C. (2016). Editorial: Linguistic influences on mathematics. *Frontiers in Psychology*, *7*, pp. 1 – 4.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effects of symbolic distance, and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, *8*(6), 612-622.

Dunn, A. L., & Tree, J. E. F. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, *12*, 273-289.

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

Ganor-Stern, D., & Tzelgov, J. (2011). Across-notation automatic processing of two-digit numbers. *Experimental Psychology*, *58*, 147 – 153.

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. In Lecleroq, P., Edmonds, A., & Hilton, *Measuring L2 proficiency: Perspectives from SLA* (pp. 208-225). UK: Multilingual Matters.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*(2), 104-122.

Goldfarb, L., Henik, A., Rubinsten, O., Block-David, Y., & Gertner, L. (2011). The numerical distance effect is task dependent. *Memory and Cognition, 39*, 1508–1517.

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

JASP Team. (2017). JASP (Version 0.8.1.2) [Computer software] 2017. *URL https://jasp-stats.org*.

Jeffreys, H. (1961). *Theory of Probability* (3rd Ed.). Oxford: Oxford University Press, Clarendon Press.

Khng, K. H., & Lee, K. (2014). The relationship between Stroop and stop-signal measures of inhibition in adolescents: influences from variations in context and measure estimation. *PloS One*, *9*(7), e101356.

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

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174.

Lim, V. P., Liow, S. J. R., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining language dominance in English-Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics*, *29*, 389–412.

Lukas, S., Krinzinger, H., Koch, I., & Willmes, K. (2014). Number representation: A question of look? The distance effect in comparison of English and Turkish number words. *Quarterly Journal of Experimental Psychology*, *67*, 260-270.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, *17*(5), 714-726.

Marsman, M. & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology, 14*(5), 545 – 555.

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

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*, 107–157.

McCloskey, M., & Macaruso, P. (1995). Representing and using numerical information. *American Psychologist, 50*, 351 – 363.

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

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual Language Development and Disorders: A handbook on bilingualism & second language learning*. Baltimore, MD: Paul H. Brookes Publishing.

Pesenti, M., & Andres, M. (2009). Common mistakes about numerical representations. Commentary on Cohen Kadosh, R. & Walsh V., Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32,* 346-347.

Pina, V., Castillo, A., Cohen Kadosh, R., & Fuentes, L. J. (2015). Intentional and automatic numerical processing as predictors of mathematical abilities in primary school children. *Frontiers in Psychology*, *6*, 375.

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

Risko, E. F., Maloney, E. A., & Fugelsang, J. A. (2013). Paying attention to attention: Evidence for an attentional contribution to the size congruity effect. *Attention, Perception, & Psychophysics*, *75*(6), 1137-1147.

Rouselle, L., & Noel, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*, 361–395.

Sasanguie, D., & Reynvoet, B. (2014). Adults' arithmetic builds on fast and automatic processing of Arabic digits: Evidence from an audiovisual matching paradigm. *PLOS One*, *9*(2), e87739.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(3), 507.

Sobel, K. V., Puri, A. M., & Faulkenberry, T. J. (2016). Bottom-up and top-down attentional contributions to the size congruity effect. *Attention, Perception, & Psychophysics*, *78*(5), 1324-1336.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315. As cited in Miller, J. O., & Pachella, R. G. (1973). Locus of the stimulus probability effect. *Journal of Experimental Psychology*, *101*(2), 227.

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

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of*

Experimental Child Psychology, *81*, 74–92.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166.

Turconi, E., Campbell, J. I., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, *98*(3), 273-285.

Wagenmakers, E.J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, A.J., … Morey, R.D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1 – 19.

Wong, B., & Szücs, D. (2013). Single-digit Arabic numbers do not automatically activate magnitude representations in adults or in children: Evidence from the symbolic same–different task. *Acta Psychologica*, *144*(3), 488-498.

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

Van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same–different task. *Journal of Cognitive Psychology*, *23*, 112-120.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66-80.

Verguts, T., & Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review*, *12*(5), 925-930.