Published in final edited form as:

Cogn Psychol. 2013 June ; 66(4): 355–379. doi:10.1016/j.cogpsych.2013.03.001.

Cross-format physical similarity effects and their implications for the numerical cognition architecture

Dale J. Cohen^{*}, Erin Warren, and Daryn Blanc-Goldhammer

Department of Psychology, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5612, United States

Abstract

The sound [faɪv] is visually depicted as a written number word “five” and as an Arabic digit “5.” Here, we present four experiments – two quantity same/different experiments and two magnitude comparison experiments – that assess whether auditory number words ([faɪv]), written number words (“five”), and Arabic digits (“5”) directly activate one another and/or their associated quantity. The quantity same/different experiments reveal that the auditory number words, written number words, and Arabic digits directly activate one another without activating their associated quantity. That is, there are cross-format physical similarity effects but no numerical distance effects. The cross-format magnitude comparison experiments reveal significant effects of both physical similarity and numerical distance. We discuss these results in relation to the architecture of numerical cognition.

Keywords

Numerical cognition; Numerical architecture; Number processing; Numerical distance effect

1. Introduction

The auditory number word [faɪv] is visually depicted as a written number word “five” and as an Arabic digit “5.” To understand these three symbols (the auditory symbol and the two visual symbols), the observer must learn the relation between each symbol and the quantity it represents, as well as the relation between the symbols themselves. To date, research in numerical cognition has primarily examined (a) whether or not all three of these symbolic formats (auditory, written, and digit) are associated with the same psychological representation of quantity (Campbell & Clark, 1988; Cohen Kadosh & Walsh, 2009; Dehaene, 1992; Gonzalez & Kolers, 1982; McCloskey, 1992; McCloskey, Caramazza, & Basili, 1985), (b) whether each symbol automatically activates its associated psychological representation of quantity (Cohen, 2009, 2010; Dehaene, 1992; Tzelgov & Ganor-Stern, 2005) and (c) the qualities of their associated psychological representation(s) of quantity (Banks & Hill, 1974; Dehaene, 1997; Dehaene, Dupoux, & Mehler, 1990; Gallistel & Gelman, 2000; Gibbon & Church, 1981; Nieder & Miller, 2003). Very few studies have addressed the interaction between these symbols. The results of such research would help clarify our understanding of the architecture of numerical cognition. For example, does the written number word “five” directly activate the Arabic digit “5,” or must a shared psychological representation of quantity mediate communication between these two symbols? Here, we present four experiments that assess whether the auditory, written, and

digit symbols denoting the integers one through nine directly activate one another and their associated quantities.

The degree to which different numerical symbols activate one another and/or their associated psychological representation of quantity is central to discovering the underlying architecture of numerical cognition. Specifically, neuroimaging data have demonstrated that numerals and the quantity information they denote are processed in separate areas of the brain (e.g., Cohen, Dhaene, Chochon, Lehericy, & Naccache, 2000; Cohen Kadosh, Cohen Kadosh, Kaas, Henik, & Goebel, 2007; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Dehaene et al., 1996; Price & Ansari, 2011). This neural modularity is the foundation of most models describing the underlying architecture of numerical cognition (e.g., Chen & Verguts, 2010; Cohen, 2010; Dehaene, 1992; McCloskey, 1992). The fundamental difference between theoretical architectures is the degree to which these modules are connected.

Fig. 1 presents several theoretical architectures of numerical cognition. The architecture at the top of the figure is a simplified depiction of McCloskey's single representation model (e.g., McCloskey & Macaruso, 1995; McCloskey, Sokol, & Goodman, 1986). In this model, each symbolic format (e.g., digit, word, etc.) is linked to the same abstract representation, but the symbolic formats are not linked to one another. Therefore, the quantity representation acts as the link between symbolic formats. For example, to convert the written number word "five" to the Arabic digit "5," the system would convert the written number word "five" into a quantity representation and then the quantity representation into the Arabic digit "5." Similarly, to compare the quantitative values of the written number word "five" and the Arabic digit "5," the system would convert both symbols into abstract quantities and then directly compare the quantity representations.

The architecture in the center of Fig. 1 is a simplified depiction of Dehaene's (1992) Triple Code Model. Similar to the McCloskey model, the Triple Code Model proposes a single abstract quantity representation to which all numerical symbolic formats are linked. However, unlike the McCloskey's model, the Triple Code Model proposes that symbolic formats are directly linked to one another. Here, the system could convert the written number word "five" directly to the Arabic digit "5," without activating the quantity representation.¹ Nonetheless, to compare the quantitative values of the written number word "five" to the Arabic digit "5," the system would convert both symbols into abstract quantities and then directly compare the quantities in the abstract quantity representation.

The architecture at the bottom of Fig. 1 is a simplified depiction of a strict Multiple Quantity Representation model (e.g., Cohen, 2009, 2010; Cohen, Ferrell, & Johnson, 2002; Gonzalez

¹Dehaene's (1992) Triple Code model assumes that the format modules link to one another and each format module links to the same abstract representation. This architecture is a version of Seidenberg and McClelland's (1989) distributed model of word recognition. Seidenberg and McClelland proposed that two format modules (orthographic and phonological coding) link to one another and to the meaning of the words. Further, the authors proposed that either the orthographic or the phonological coding can activate the meaning of the word, with the orthographic coding being the primary pathway to meaning in reading. Nevertheless, this model assumes that all pathways between format modules and meaning can be activated in parallel. The model does not, unfortunately, specify a priori when each module will be activated and used. As such, this model does not make precise predictions to be assessed in the present set of experiments. To permit precise predictions, we assume each module activated plays a role in processing the stimuli. We do not consider the predictions of models that assume activations of modules that play no part in the processing of stimuli (though they are discussed in Section 6). For example, Seidenberg and McClelland's (1989) model can be interpreted as predicting that all modules will be activated all the time, regardless of whether the modules play an active part in recovering meaning. Such a model would predict we would find evidence for the activation of all the different modules (every feature module and the quantity representation) in every condition. Although this incantation of the Seidenberg and McClelland's (1989) model can produce a precise prediction, other versions do not. For example, one can imagine countless versions in which some, but not all, modules are activated and a subset of those are involved in the processing of the stimuli. Because such "in between" versions can accommodate all patterns of data, they will not be considered here. Rather, we do not consider models in which modules are activated but play no role in processing the stimuli.

& Kolers, 1982). Here, each numerical symbolic format activates a separate quantity representation, but symbolic formats are directly linked to one another. The Multiple Quantity Representation model is a particular instantiation of theories that discuss a non-abstract quantity representation (e.g., Cohen Kadosh & Walsh, 2009). These theories do not assume that the “neural populations that code numerical quantity are insensitive to the form of input in which the numerical information was presented” (Cohen Kadosh & Walsh, 2009, p. 314). Similar to the Triple Code Model, this system is able to convert the written number word “five” directly to the Arabic digit “5,” without activating a quantity representation. Unlike both previous models, to compare quantities of the written number word “five” to the Arabic digit “5,” the system must first convert one symbol (e.g., “five”) into the format of the other symbol (e.g., “5”). Once converted into the same format, the two numerals would activate the same quantity representation and their quantities could be efficiently compared.²

As described above, the fundamental difference between the proposed architectures is not their modules, but rather the links between the modules. The links between the modules proposed by different architectures of numerical cognition can be discovered, at least in part, by assessing the activation of each module in different tasks. Traditionally, the *numerical distance effect* has been used as a signature for the activation of the quantity module associated with a numerical symbol. The numerical distance effect refers to the pattern of data whereby integers that are quantitatively close interfere with one another more than integers that are quantitatively distant. This interference results from overlapping perceptual distributions of quantity (e.g., Buckley & Gillman, 1974; Dehaene, 1997; Gallistel & Gelman, 1992). In tasks that require quantity comparison (e.g., judging the larger of two stimuli), overlapping perceptual distributions of quantity increase the perceived similarity between the stimuli, resulting in slower reaction times (RT). In this instance, the numerical distance effect is well described by the Welford (1960) function:

$$RT = a + k * \log [L / (L - S)] \quad (1)$$

where a and k are constants, L is the larger quantity in the comparison, and S is the smaller quantity. In tasks that do not require quantity comparison (e.g., judging the sameness of two stimuli), overlapping perceptual distributions of quantity interfere with the identification of the stimulus, and again result in slower RTs. When the numerical distance effect is present in a task that does not require the quantity of the stimulus to be assessed, researchers typically conclude that the quantity representation was automatically activated (e.g., Dehaene & Akhavein, 1995; Ganor-Stern, Tzelgov, & Ellenbogen, 2007; Tzelgov & Ganor-Stern, 2005).

Recently, Cohen (2009) identified a signature pattern of data for the activation of the Arabic digit symbolic format based on the physical structure of the numeral. To assess whether the physical structure of an Arabic digit influenced RT, Cohen (2009) developed a *Physical Similarity* function of Arabic digits (PS_{digit}) based on the seven line, Fig. 8 structure of numbers used on digital alarm clocks:

$$PS_{\text{digit}} = O / D \quad (2)$$

where PS_{digit} is the measure of physical similarity, O is the number of lines that the two integers share, and D is the number of remaining lines (see Table 1). This function allows for an objective quantification of the physical similarities between integers. Cohen (2009)

²We have some evidence that comparisons between quantity representation modules are possible (Warren & Cohen, 2013). However, because such comparisons appear to be both noisy and slow, and the evidence is relatively new, we do not consider it. Here, we are testing the strict Multiple Representation model in which there is no direct communication between quantity representations.

assessed whether physical similarity or numerical distance was the primary influence on RT in a quantity same/different task. The participants were presented with the integers 1–9, in Arial font, and were instructed to judge whether the Arabic digit presented was a 5. Cohen found that the PS_{digit} function was the primary predictor of the RT data. To ensure the results were not due to a unique characteristic of Arial font, Cohen (2009) completed a second experiment in which the font of the numerical stimuli was randomized. In Experiment 2, Cohen again found that physical similarity was the primary determinant to the participants' RTs.

Cohen's (2009) PS_{digit} function can be used as a signature for the activation of the symbolic digit associated with another numerical symbolic format or a quantity. Presumably, the PS_{digit} function results because overlapping perceptual distributions of physical structure interfere with the identification of the stimulus, and this interference slows RTs. As such, the presence of the PS_{digit} function in a task suggests that physical structure of a digit was activated. If the PS_{digit} function is present in a task whereby participants are not required to activate the interfering Arabic digit, one can conclude that the interfering digit automatically activated the Arabic digit module. Thus, with the availability of the numerical distance effect and physical similarity function, one can assess if and when a numeral's quantity and symbolic representations are automatically activated.

To assess the activation of the symbolic written number word and the symbolic auditory number word, we introduce two new physical similarity functions: PS_{word} and PS_{auditory} (see Table 1). These functions are not intended to be definitive descriptions of the perceptual similarities between written number words or auditory number words. Rather, these functions are intended to capture the physical features sufficient to reveal the influence of physical similarity in number judgment tasks.³

We based PS_{word} on two features of the written number words to be compared: similarity in length and the visual confusion of the letters. We used Geyer's (1977) confusion matrix for lowercase alphabetic characters as the values of the letter confusions: higher numbers indicate greater confusion, thus more similarities between the two letters. The letter confusion value equaled the average confusion of: the first letter of the two words, the last letter of the two words, the remaining letters of the two words, and all the letters of the two words. We included the first (LC_{first}) and last letters (LC_{last}) in the formula because of their relative importance in reading a word (e.g., McCusker, Gough, & Bias, 1981). To capture the intraword confusions, we included $LC_{\text{remaining}}$. Finally, we included a measure of all confusions (LC_{all}) to capture any global effects. So, the letter confusions of Word A and Word B (LC_{AB}) is calculated by,

$$LC_{AB} = (LC_{\text{first}} + LC_{\text{last}} + LC_{\text{remaining}} + LC_{\text{all}}) / 4 \quad (3)$$

The similarity in word length equaled the absolute value of the number of letters in Word A minus the number of letters in Word B. So, the word length confusion of Word A and Word B (WL_{AB}) is calculated by,

$$WL_{AB} = \text{abs}(WL_A - WL_B) \quad (4)$$

³We assume the numerical cognition system's comparison process involves comparing the some aspect of the physical features of the numerical symbols (once both symbols have been converted into the same format). As such, we assume that the process relates to physical similarity. It is likely that our formulas do not capture the comparison process perfectly. Therefore, the delta between our measures of physical similarity and the actual process will likely result in the smaller effect sizes. Nevertheless, because our measures of similarity capture some important aspects of the process, they should show an effect if it is present.

Finally, PS_{word} is the average of these two values,

$$PS_{\text{word}} = (LC_{AB} + WL_{AB}) / 2 \quad (5)$$

Fig. 2 illustrates the similarity of the number words “two” and “five.” These two words have relatively high similarities on our PS_{word} measure. As the figure shows, both words are of fairly similar length. In addition, the first and last letters of both words have a relatively similar shape. The remaining letters were somewhat similar, especially the “w” in two and the “v” in five.

We used Shepard, Kilpatrick and Cunningham’s (1975) MDS analysis of the spoken number word to generate ordinal similarity distances for PS_{auditory} . Thus, we did not identify features per se. Rather, we based our similarity assessment on the MDS distances provided. Nevertheless, one can make an educated guess. To illustrate, the similarities of the spoken number words [faɪv] (five) and [eɪt] (eight), likely revolves around the fact that both words are dominated by a diphthong that ends in the [i] sound. We identified these distances only for the number pairs present in the current experiment and present them in Table 1. Table 2 presents the correlations between the physical similarity functions and the Welford function.

Each of the models of the architectures underlying numerical cognition described above makes different predictions concerning the activation of the symbol and quantity modules in cross-format magnitude comparison task and cross-format quantity same/different task (see Table 3). Here, we assume that an effect of numerical distance is a marker for the activation of the quantity module and an effect of physical similarity is a marker for the activation of the symbolic format module by both numeric symbols.⁴ Furthermore, we assume that the quantity module(s) and/or the symbolic format modules will not be activated unless the architecture of the numerical cognition model necessitates their activation. Finally, we assume the most direct route for completion of the task.

The quantity same/different task requires the participant to judge whether two presented numerals denote the same quantity. Theoretically, when the two numerals are presented in the same-format, this task can be completed simply by analyzing the surface structure of the presented numerical symbols (Cohen, 2009; e.g., same: 5 vs. 5; different: 5 vs. 6). This is because the surface structure of the symbols are perfectly correlated with the quantity they denote and when the two numerals are presented in the same-format, then their surface structure will be identical if and only if they denote the same quantity. The cross-format quantity same/different task, in contrast, cannot be completed by a surface structure analysis. This is because the surface structure of the two presented numerals will not be identical whether they denote the same or different quantities. In the cross-format condition, the architecture of the numerical cognition system becomes crucial for predicting the outcome of this task. Suppose two number symbols are presented in two different formats, Format A (e.g., “five”) and Format B (e.g., “5”). Here, the surface structures of the presented symbols are different. If, however, the symbolic format modules of the two presented number symbols are directly linked, then the number symbol in Format A (“five”) can *activate* its counterpart in Format B (“5”). Once the number symbol in Format A has activated its counterpart in Format B, the two activated symbols in Format B (e.g., “5” and “5”) can be compared based on non-quantitative information (e.g., surface structure). If the architecture

⁴Here, when we state that a module is activated, whether a quantity module or a symbolic format module, we are indicating the module is activated by *both* integers. That is, when the quantity module is activated, interference between the quantities of the two presented numbers results in slower RTs. When the symbolic format module is activated, interference between the physical similarities of the two presented numbers results in slower RTs. It is possible that these modules may be activated by a single quantity or a single symbol. However, such activation will not result in either the numerical distance effect or the PS effect.

of the numerical cognition system permits such a conversion and comparison, one would expect to see RT predicted by Physical Similarity function but not the Numerical Distance function. This is what the Triple Code and Multiple Representation Models predict for the cross-format quantity same/different task. In contrast, if the symbolic format modules are independent (as is the case for the McCloskey model, see McCloskey & Macaruso, 1995, p. 357), then the only available route for comparison in the cross-format same/different task is through the quantity representation. In such an instance, each numerical symbol activates its' quantity in the shared quantity representation and the quantities are compared, rather than the symbol's physical structure. Here, one would expect to see RT predicted by Numerical Distance but not Physical Similarity. This is what McCloskey's model predicts for the cross-format quantity same/different task.

In contrast to the quantity same/different task, the magnitude comparison task presumably cannot be completed simply by analyzing the surface structure of the presented numeral whether or not the numerals are presented in the same format. Because two different numerals are presented on every trial, their surface structures are different regardless of which presented numeral is greater. In addition, for the magnitude comparison task to be completed simply by analyzing the surface structure of the numerical symbols, one would have to hypothesize a non-analogue representation of quantity.⁵ Where as there exists some evidence for this when assessing decimals (e.g., Cohen, 2010), to date there exists no evidence for a non-analogue representation of quantity for integers.

Presumably the numerical symbol's quantitative meaning must be accessed in all conditions of the magnitude comparison task. Therefore, all models of the architecture of numerical cognition predict that a quantity module will be activated, resulting in an effect of numerical distance. Here, we describe the predictions for the cross-format magnitude comparison task because this is where the models' predictions differ. If the modeled architecture specifies that the symbolic format modules share a single, abstract quantity representation, then quantity comparison can occur within that shared representation, regardless of the formats of the numerical symbols. That is, each numerical symbol will activate its' quantity in the shared quantity representation and the quantities will be compared. In this instance, one would expect to see RT predicted by Numerical Distance but not Physical Similarity. This is what the Triple Code and McCloskey Models predict for the cross-format magnitude comparison task. If, however, the modeled architecture specifies that the symbolic format modules do not share a quantity representation, then magnitude comparison is not as straight forward. Here, the two numerical symbols must first be converted into the same format before magnitude comparison can occur. This is because quantity comparison can only occur in a shared quantity representation, and the two presented numerals will only share a quantity representation when they are in the same format. Once both symbols are in the same format, magnitude comparison can occur within the single quantity representation. Here, because both the symbolic format module and the quantity module will be activated one would expect to see RT predicted by both Physical Similarity and Numerical Distance. This is what the Multiple Representation Model predicts for the cross-format magnitude comparison task.

There has been no published research addressing the presence of the PS and Welford functions together in a cross-format magnitude comparison task. Therefore, the current literature cannot address the activation of the numerical symbolic format module in a magnitude comparison task. There is, however, abundant evidence that both cross-format

⁵If (a) the magnitude comparison task requires knowledge of quantity, (b) the magnitude comparison task is completed successfully, and (c) no analogue quantity representation is activated, then it logically follows that a non-analogue quantity representation is available.

and same format magnitude comparison tasks elicit the numerical distance effect (e.g., Buckley & Gillman, 1974; Dehaene, 1989; Dehaene & Akhavein, 1995; Dehaene, Dupoux & Mehler, 1990; Duncan & McFarland, 1980; Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1967, 1973; Shepard et al., 1975; Takahashi & Green, 1983).

Similarly, there has been no published research addressing the presence of the PS and Welford functions together in a cross-format quantity same/different task. Nevertheless, there have been several studies that examine the presence of the numerical distance effect in a cross-format quantity same/different task. Recall, the numerical distance effect has been used as evidence that numerical symbols automatically activate their associated quantity representations. Specifically, if a symbol's quantity interferes in a task in which quantity is irrelevant, researchers conclude that the quantity representation was automatically activated (for a review, see Tzelgov & Ganor-Stern, 2005). The quantity same/different task is one such task.

Dehaene and Akhavein (1995) completed a cross-format quantity same/different task examining Arabic digits and written number words. In one experiment, the authors had participants compare either two Arabic digits, two written number words (e.g. TWO), or a mixed trial of one Arabic digit and one written number word. The numbers used were 1, 2, 8, and 9 and the participants were asked to determine whether the two quantities were the same or different. To analyze the data, the authors dichotomized the number pairs into "numerically close" (i.e., 1 vs. 2 and 8 vs. 9) and "numerically far" (1 or 2 vs. 8 or 9) conditions. The authors found that participants responded more quickly to the numerically far condition than the numerically close condition in all format combinations. The authors also found that participants responded more slowly to cross-format trials than same format trials – even when the cross-format stimuli represented the same quantity. In a second experiment, the authors conducted a form same/different task, whereby participants had to judge whether the physical form of the presented stimuli were the same (e.g. respond same when, ONE ONE; respond different when, ONE 1). Although participants responded more quickly to the numerically far stimuli than the numerically close stimuli in the same format conditions, quantity did not affect the RTs in the cross-format conditions. Disregarding this latter finding, the authors concluded that the majority of the evidence suggests that numerical symbols automatically activated their quantity representation. Ganor-Stern and Tzelgov (2008) conducted an analogous experiment using Indian and Arabic notation and found similar results. In both studies, the influence of physical similarity was not assessed and thus not known.

In contrast to the above data, Cohen (2009) revealed that quantity representations are not automatically activated in a quantity same/different task. As discussed above, Cohen (2009) showed that the PS function accounted for virtually all the variance in the RTs in a same format quantity same/different task. Indeed, because the pattern of data resulting from the PS function is often similar to that resulting from the Welford function, those studies that do not assess the influence of both functions may mistake effects of physical similarity for those of numerical distance. Because significant collinearity between the Welford function and physical similarity function is also evident in the digit and auditory PS functions (see Table 2), researchers may make similar mistakes when assessing cross-format distance effects.

The influence of physical similarity on the data in a cross-format quantity same/different task remains unclear. For example, suppose the standard is presented in a word format and the probe in a digit format. Here, there is no surface level physical similarity between the two presented numerals. Therefore, if the PS function is primarily a result of a visual comparison of features, the PS function should not manifest in such a task. However, if the

PS function results from the specifics of the architecture of the numerical cognition system, then the PS function may manifest in the data (e.g., the Multiple Representation Model). To clarify, the architecture of the numerical cognition system may have evolved so the most efficient comparison of sameness of two symbols occurs by converting one symbol into the format of the other symbol. Here for example, such a system may convert the standard into the format of the probe prior to comparison. After such a conversion, the surface structure of the two symbols can be compared and thus the PS function would manifest. Without the format conversion, there is little a priori reason for the PS function to manifest. Furthermore, without the format conversion, the numerical distance function would likely manifest because both symbols would likely be converted into a quantity prior to comparison.

In the present experiments, we (a) assess the underlying architecture of numerical cognition in a series of quantity same/different and magnitude comparison tasks, and (b) expand on Cohen's (2009) findings by having participants compare Arabic digits to the quantities 2, 5 and 8 (rather than just 5).

2. Experiment 1

In Experiment 1, we assessed the relative influence of the PS and Welford functions in a quantity same/different task. The numerals were presented as written number words (e.g., two) or Arabic digits (e.g., "2"). We assessed both same-format and cross-format conditions.

2.1. Method

2.1.1. Participants—Seventy naïve undergraduate volunteers participated for class credit.

2.1.2. Apparatus and stimuli—All the stimuli were presented on a 24-in. LED color monitor with a 72-Hz refresh rate controlled by a Macintosh Mini running an OS X. The resolution of the monitor was 1920 × 1200 pixels.

A trial consisted of either two Arabic digits, two written number words, or an Arabic digit and a written number word, presented one above the other, in the center of the screen. The stimuli were presented vertically (one above the other) rather than horizontally (one to the right of the other) to minimize the less relevant influence of spatial information (e.g., Dehaene, Bossini, & Giraux, 1993). The standard values were 2, 5, and 8 and were always presented above the probe. The probe values were selected randomly from 1 to 9. All the Arabic digits were presented in Arial font.

2.1.3. Procedure—The participants were tested individually in a small, dark room. The participants were instructed that, on every trial, they would be presented with two quantities, one above the other, and that they were to determine whether the bottom quantity was the same or different than the top quantity. Half the participants were told to press the "D" key if the two quantities were the same and the "K" key if the two quantities were different. The keys were reversed for the remaining participants. The participants were instructed that speed was important but that accuracy was essential.

Each trial consisted of the presentation of a standard and probe and then the participant's response. The stimuli remained on the screen until the participant responded. The computer recorded the participant's RT. There was a 500-ms delay between trials.

There were four blocked conditions: two same format and two cross-format. The same format conditions consisted of: a written number word as the standard and a written number word as the probe (word–word); an Arabic digit as the standard and an Arabic digit as the probe (digit–digit). The cross-format conditions consisted of: a written number word as the

standard and an Arabic digit as the probe (word–digit); an Arabic digit as the standard and a written number word as the probe (digit–word). Each condition consisted of 288 trials. Half the trials had a correct response of “same” and half were “different.” There were 8 “different” probes per standard (the numbers 1–9 excluding the presented standard) and 3 standards (the numbers 2, 5, and 8). The 288 trials consisted of 3 (standards) $\times$ 8 (probes) $\times$ 2 (same/different) $\times$ 6 (repetitions).

Conditions were blocked and the 288 trials were presented randomly within each block. Each block began with 12 practice trials. After every block of trials, the experiment would pause and ask the participant whether he or she would like to take a self-timed break. The experiment resumed when the participant pressed the return key. Each of the four blocked conditions lasted about 15 min. The order of the conditions was counterbalanced across participants.

2.2. Results

Prior to analyzing the data, we (a) normalized the RT data using a log transformation, (b) trimmed the fastest and slowest 0.5% of the responses, and (c) removed all incorrect responses (average error rate of 4%). All together, about 5% of the data was removed. Finally, we removed four participants whose error rates exceeded 10.5%. To keep the analyses consistent across models, only the trials in which the standard and the probe denoted different quantities (i.e., the “different” trials) were analyzed. Table 4 presents the means (sds) and error rates for each condition.

To assess the relative contributions of the Welford and the Physical Similarity (PS) functions, for each condition, we computed a simultaneous Mixed Model regression with RT for each probe as the criterion variable, the Welford and Physical Similarity functions as the predictor variables, and subject as a random variable (function *lmer* in R). To assess the relative contribution of each predictor, we standardized all variables. To provide a measure of variance accounted for, we report R^2 , calculated by squaring the correlation between raw RT and the fitted values. Because the current Mixed Model analysis fits functions to each individual’s raw data, rather than the mean data, these values appear smaller than those found in the extant literature. Generally, when reporting R^2 for the numerical distance effect, researchers report the r^2 for the mean RTs, collapsing across trials and participants. By removing individual trial variance, researchers can attain r^2 values of 0.8 and higher. Here, we are fitting to all the raw data and we attain R^2 values of around 0.2 and greater (which correspond to medium and large effect sizes). When we collapse across trials and participants, our R^2 values are comparable to those in the extant literature. Because each analysis assessed four slopes for significance, we used the Bonferroni correction, requiring an alpha of 0.0125. Fig. 3 shows boxplots of the participants’ slopes for each predictor by condition. Table 5 summarizes the significant predictors for each condition. Table 6 presents the slopes and t -values of all predictors.

When two Arabic digits were presented (digit–digit), the slope for PS_{digit} (slope = 0.09, $t = 9.2$, $p < .01$), PS_{auditory} (slope = 0.05, $t = 4.2$, $p < .01$) and the intercept (-0.36 , $t = -6.9$, $p < .01$) were significant⁶ ($R^2 = .22$). There were no other significant predictors.

When a written number word was the standard and an Arabic digit was the probe (word–digit), the slope for PS_{digit} (slope = 0.07, $t = 6.3$, $p < .01$) and the intercept (0.25 , $t = -4.2$, $p < .01$) were significant ($R^2 = .25$). There were no other significant predictors.

⁶Because we standardized all variables, a significant slope roughly represents the proportion of an SD change in RT that is predicted by 1 SD change in the predictor variable.

When two written number words were presented (word–word), the slope for the PS_{word} (slope = 0.07, $t = 7.5$, $p < .01$) and the intercept (-0.18 , $t = -3.4$, $p < .01$) were significant ($R^2 = .22$). There were no other significant predictors.

When an Arabic digit was the standard and a written number word was the probe (digit–word), the slope for the PS_{auditory} (slope = 0.05, $t = 4.3$, $p < .01$) and the intercept (0.28 , $t = 5.3$, $p < .01$) were significant ($R^2 = .21$). There were no other significant predictors.

2.3. Discussion

The data are clear, in all conditions participants completed the same/different task using the physical similarity of the standard to the probe. The non-quantitative route was used in both the same-format and cross-format conditions. This is the first time a cross-format physical similarity effect has been demonstrated to be the primary controlling variable in a quantity same/different task. Although in most cases, physical similarity was based on the format of the probe stimulus, the auditory similarity was a significant predictor in several conditions.

The significant effects of physical similarity indicate that there is cross activation of numerical formats. Because there was no effect of numerical distance, an abstract quantity representation could not have been the route for the cross activation to have occurred. These data confirm and extend the results of Cohen (2009) beyond the standard of 5. Together these results suggest that there exists a direct link between the number word symbolic format module and the Arabic digit symbolic module.

We discuss the findings of Experiment 1 in more detail in Section 6. In Experiment 2, we conduct a traditional magnitude comparison task to further explore the links between the Arabic digit symbolic format, written number word symbolic format, and quantity representation modules.

3. Experiment 2

Experiment 2 was designed to expand on the findings of the quantity same/different task of Experiment 1. Here, we conduct a classic magnitude comparison experiment using the same parameters as Experiment 1. If the Arabic digit symbolic format module and the written number word symbolic format module share a quantity representation, then there should be a significant effect of the Welford function and no effect of the PS functions in the cross-format conditions. Cohen (2010) found an effect of physical similarity with two Arabic digits, using the classic magnitude comparison task. Here, we expand on these findings using single Arabic digits and written number words.

3.1. Method

3.1.1. Participants—Seventy naïve undergraduate volunteers participated for class credit.

3.1.2. Apparatus and stimuli—The apparatus and stimuli used in Experiment 2 were the same as Experiment 1.

3.1.3. Procedure—The procedure for Experiment 2 was identical to Experiment 1 with the following exceptions. The participants were instructed that, on every trial, they would be presented with two quantities, one above the other, and that they were to determine whether the bottom quantity was larger or smaller than the top quantity. Furthermore, there were no trials in which the standard was the same quantity as the probe. The 288 trials consisted of 3 (standards) × 8 (probes) × 12 (repetitions).

3.2. Results

Similar to Experiment 1, prior to analyzing the data, we (a) normalized the RT data using a log transformation, (b) trimmed the fastest and slowest 0.5% of the responses, and (c) removed all incorrect responses (average error rate of 5.4%). All together, about 6.4% of the data was removed. Finally, we removed 13 participants whose error rates exceeded 10.5%. Table 4 presents the means (sds) and error rates for each condition.

Similar to Experiment 1, we computed a simultaneous Mixed Model regression with RT for each probe as the criterion variable, the Welford and Physical Similarity functions as the predictor variables, and subject as a random variable. To assess the relative contribution of each predictor, we standardized all variables and required an alpha of 0.0125. Fig. 4 shows boxplots of the participants' slopes for each predictor by condition. Table 5 summarizes the significant predictors for each condition. Table 6 presents the slopes and t -values of all predictors. When two digits were presented (digit–digit), the slope for PS_{digit} (slope = 0.04, $t = 6.1$, $p < .01$), the Welford function (slope = 0.1, $t = 8.7$, $p < .01$) and the intercept (-0.41 , $t = -5.8$, $p < .01$) were significant ($R^2 = .33$). There were no other significant predictors.

When a number word was the standard and a digit was the probe (word–digit), the slope for PS_{digit} (slope = 0.05, $t = 4.1$, $p < .01$), PS_{word} (slope = 0.03, $t = 3.5$, $p < .01$), and the Welford function (slope = 0.12, $t = 13$, $p < .01$) were significant ($R^2 = .30$). There were no other significant predictors.

When two words were presented (word–word), the slope for PS_{word} (slope = 0.04, $t = 5.3$, $p < .01$), the Welford function (slope = 0.14, $t = 12.5$, $p < .01$) and the intercept (0.23, $t = 3.3$, $p < .01$) were significant ($R^2 = .31$). There were no other significant predictors.

When an Arabic digit was the standard and a written number word was the probe (digit–word), the slope for the Welford function (slope = 0.10, $t = 8.7$, $p < .01$) was significant ($R^2 = .29$). There were no other significant predictors.

3.3. Discussion

The data show that the RTs of each of the four conditions were best described by the Welford function. This suggests that numerical distance is the primary controlling factor in the participants' magnitude comparison process. Such a finding merely confirms the decades of research that demonstrates the activation of the quantity representation module in the magnitude comparison task. However, the data also indicate that in three of the four conditions, there was also an effect of physical similarity. This finding is new and demonstrates the activation of complementary symbolic format modules.

The cross-format conditions are somewhat contradictory with regards to the predictions in Table 3. Specifically, the results of the word–digit condition are consistent with the Multiple Representation Model (i.e., effects of both numerical distance and physical similarity), but the results of the digit–word condition are consistent with the Triple Code Model (i.e., only an effect of numerical distance). We discuss these results in detail in Section 6.

In Experiments 3 and 4, we expand our findings to the auditory symbolic format. Specifically, in Experiment 3, we assess the auditory number words and Arabic digits in a quantity same/different task. In Experiment 4, we assess the auditory number words and Arabic digits in a magnitude comparison task. Both experiments are very similar to Experiments 1 and 2, with the exceptions that (a) the standard and the probes are presented successively rather than simultaneously, and (b) we only assessed the Arabic digit probe conditions. We presented the standard and the probe successively because the auditory presentation of a number was distracting when presented simultaneously with a visual

stimulus. Because the auditory stimulus is presented in time and then is gone, participants must pay attention to the auditory stimulus when it is presented for optimal performance. We only assessed the digit probe condition because, at the time we designed the experiment, we did not have a PS function for auditory number words and we (incorrectly) assumed the standard would always be converted into the format of the probe. Therefore we did not anticipate assessing the physical similarity effect in the auditory probe condition (which are the conditions in which we would expect an auditory physical similarity effect).

4. Experiment 3

Experiment 3 was designed to expand on the findings of Experiments 1 and 2 using Arabic digits and auditory number words as stimuli. Here, we conduct a quantity same/different task using similar parameters as Experiment 1.

4.1. Method

4.1.1. Participants—Fifty-nine naïve undergraduate volunteers participated for class credit.

4.1.2. Apparatus and stimuli—In addition to the apparatus and stimuli used in Experiment 1, we used verbal recordings of a female voice reading the standard numbers 2, 5, and 8. These recordings were sampled at 44.1 kHz. The sound files were trimmed to remove excess time before the start and after the end of the number word. Each recorded number lasted about 500 ms ($|tu| = 411$ ms, $|fav| = 475$ ms, $|eit| = 498$ ms). The participants wore Sennheiser HD 280 Pro headphones in all conditions (i.e., whether or not audio stimuli were playing).

4.1.3. Procedure—The procedure for Experiment 3 was identical to Experiment 1 with the following exceptions. In both the digit–digit and the auditory–digit conditions, the standard and the probe were presented successively. The inter-stimulus interval (ISI) between the standard and probe was 500 ms. In the digit–digit condition, the standard was presented with a slight blue tint to differentiate the standard from the probe (and thus cue the participant which stimulus to respond to). The probe was a light gray (as it was in all experiments). All participants wore headphones.

4.2. Results

Similar to Experiment 1, prior to analyzing the data, we (a) normalized the RT data using a log transformation, (b) trimmed the fastest and slowest 0.5% of the responses, and (c) removed all incorrect responses (average error rate of 3.3%). All together, about 4% of the data was removed. Finally, we removed three participants whose error rates exceeded 10.5%. To keep the analyses consistent across models, only different trials were analyzed. Table 4 presents the means (sds) and error rates for each condition.

Similar to Experiment 1, we computed a simultaneous Mixed Model regression with RT for each probe as the criterion variable, the Welford and Physical Similarity functions as the predictor variables, and subject as a random variable. To assess the relative contribution of each predictor, we standardized all variables and required an alpha of 0.0125. Fig. 5 shows boxplots of the participants' slopes for each predictor by condition. Table 5 summarizes the significant predictors for each condition. Table 6 presents the slopes and t -values of all predictors.

When two Arabic digits were presented (digit–digit), the slope for PS_{digit} (slope = 0.05, $t = 4.9$, $p < .01$) was significant ($R^2 = .28$). There were no other significant predictors.

When an auditory number word was the standard and an Arabic digit was the probe (auditory–digit), the slope for PS_{digit} (slope = 0.04, $t = 2.5$, $p = .012$) was significant ($R^2 = .28$). There were no other significant predictors.

4.3. Discussion

The data are clear, participants completed the quantity same/different task using a non-quantitative route based on the physical similarity of the standard to the probe in both the digit–digit condition and the auditory–digit condition. Interestingly, the PS_{auditory} function was not found significant, in the digit–digit condition, as it was in Experiment 1. These data extend the findings of Experiment 1 to the auditory domain (i.e., cross-modality). We discuss the findings of Experiment 3 in more detail in Section 6.

In Experiment 4, we conduct a traditional magnitude comparison task to further explore the links between the Arabic digit symbolic format, auditory number word symbolic format, and quantity representation modules.

5. Experiment 4

Experiment 4 was designed to expand on the findings of the quantity same/different task of Experiment 3 using Arabic digits and auditory number words as stimuli. Here, we conduct a magnitude comparison experiment using the same parameters as Experiment 3. If the Arabic digit symbolic format module and the auditory number word symbolic format module share a quantity representation, then there should be a significant effect of the Welford function and no effect of the PS functions in the cross-format conditions.

5.1. Method

5.1.1. Participants—Fifty-eight naïve undergraduate volunteers participated for class credit.

5.1.2. Apparatus and stimuli—The apparatus and stimuli used in Experiment 4 were the same as Experiment 3.

5.1.3. Procedure—The procedure for Experiment 4 was identical to Experiment 3 with the following exceptions. The participants were instructed that, on every trial, they would be presented with two quantities, one before the other, and that they were to determine whether the second quantity was larger or smaller than the first quantity. Furthermore, there were no trials in which the standard was the same quantity as the probe. The 288 trials consisted of 3 (standards) × 8 (probes) × 12 (repetitions).

5.2. Results

Similar to Experiment 1, prior to analyzing the data, we (a) normalized the RT data using a log transformation, (b) trimmed the fastest and slowest 0.5% of the responses, and (c) removed all incorrect responses (average error rate of 5.4%). All together, about 6.4% of the data was removed. Finally, we removed three participants whose error rates exceeded 10.5%. Table 4 presents the means (sds) and error rates for each condition.

Similar to Experiment 1, we computed a simultaneous Mixed Model regression with RT for each probe as the criterion variable, the Welford and Physical Similarity functions as the predictor variables, and subject as a random variable. To assess the relative contribution of each predictor, we standardized all variables and required an alpha of 0.0125. Fig. 6 shows boxplots of the participants' slopes for each predictor by condition. Table 5 summarizes the

significant predictors for each condition. Table 6 presents the slopes and t -values of all predictors.

When two Arabic digits were presented (digit–digit), only the Welford function (slope = 0.03, $t = 2.5$, $p = .012$) was significant, although PS_{auditory} approached significance ($t = 2.4$) ($R^2 = .36$). There were no other significant predictors.

When an auditory number word was the standard and an Arabic digit was the probe (auditory–digit), the slope for PS_{auditory} (slope = 0.04, $t = 2.9$, $p < .01$) and the Welford function (slope = 0.07, $t = 4.8$, $p < .01$) were significant ($R^2 = .30$). There were no other significant predictors.

5.3. Discussion

The data show that the Welford function was the primary controlling factor when determining whether the standard was the same as the probe for both the digit–digit and the auditory–digit conditions. In both conditions, there was a strong influence of physical similarity based on the spoken number words, however, in the digit–digit condition, the PS_{digit} function did not reach significance. We discuss the findings of Experiment 4 in more detail in Section 6.

6. General discussion

The data from Experiments 1–4 provide some insight into the architecture of the numerical cognition system. In all conditions in the quantity same/different task, the PS function was the primary controlling factor and there were no effects of numerical distance. This is the first time cross-format physical similarity effects have been demonstrated. This clearly reveals that observers make quantity same/different judgments based on non-quantitative information, even in cross-format conditions. In all conditions of the magnitude comparison task, the Welford function was significant. Furthermore, in all but one condition of the magnitude comparison task, the PS function was also significant. Below we interpret these results in the context of the automatic activation of symbol and quantity modules and the architecture of numerical cognition. First, however, we address why our results differ from other published results studying cross-format same/different numerical judgments (Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008; Van Opstal & Verguts, 2011; Verguts & Van Opstal, 2005).

6.1. Cross-format same/different numerical judgment

In the present quantity same/different experiments, we found that physical similarity was the primary predictor of RTs, rather than the magnitude that the symbols denoted. This finding was true both in the same-format and the cross-format conditions. Previous studies addressing this issue have reported numerical distance effects when assessing numerical symbols in both same-format and cross-format same/different tasks. The questions naturally arises, “why does our data differ from the other published data?” We address that question here.

Dehaene and Akhavein (1995) appeared to have set the standard for designing the typical numerical same/ different task. The authors used the integers 1, 2, 8, and 9, and classified the integers as either quantitatively close (e.g., 1 2) or quantitatively distant (e.g., 1 8). This categorization reduces the strong predictions of the continuous numerical distance function into a simple dichotomous categorization. This reduction eliminates much of the power to distinguish the numerical distance effect from other effects that are correlated with it. Cohen (2009) showed that physical similarity is strongly correlated with the numerical distance effect. García-Orza, Perea, Mallouh, and Carreiras (2012) found a similarly strong

correlation between numerical distance and the physical similarity of Persian–Indian integers. Thus, the use of few numerical integers, the classification of these integers into a small number of distance categories, and the strong correlation between physical similarity and numerical distance created a scenario whereby RTs primarily produced by physical similarity appeared to be consistent with the numerical distance effect. It is only when a large number of numerical stimuli are assessed (not categorized) and the numerical distance and physical similarity are pitted against one and other that the two competing hypotheses can be distinguished. This is exactly what Cohen (2009) did in a same-format same/different task when he demonstrated the primary influence of physical similarity rather than numerical distance. Since that time, several other researchers have confirmed Cohen's (2009) effect for same-format stimuli under a variety of conditions (García-Orza et al., 2012; Goldfarb, Henik, Rubinsten, Bloch–David, & Gertner, 2011). Indeed, Goldfarb et al. (2011) demonstrated that the numerical distance effect is evident when the number stimuli are categorized similar to Dehaene and Akhavein (1995), but the numerical distance effect disappears when the number stimuli are uncategorized (Verguts & Van Opstal, 2005, also noted the physical similarity of 8 and 9 when discussing Dehaene & Akhavein, 1995).

The studies that have demonstrated the numerical distance effect in the published literature have all used the Dehaene and Akhavein (1995) categorization analysis and have not assessed the influence of physical similarity (e.g., Dehaene & Akhavein, 1995; Ganor-Stern & Tzelgov, 2008; Van Opstal & Verguts, 2011; Verguts & Van Opstal, 2005). As such, the published studies have not provided an adequate test of the physical similarity hypothesis. Only Van Opstal and Verguts (2011) decomposed their four numerical stimuli (1, 2, 7, & 8) into numerical distances (in addition to their categorical analysis). The authors found a significant effect of numerical distance in their cross-format condition (digit–word). However, the authors only assessed four numerical distances (1, 5, 6, & 7) and never assessed the effects of physical similarity. Importantly, the numerical distance effect emerges in our data in the cross-format digit–word and word–digit conditions when we assess only the effect of numerical distance (without including the physical similarity variables). That is, we find a significant effect of the Welford function ($t_s > 2.7$, $p < .01$). Importantly, this effect emerges even though we assessed far more numerical distances than Van Opstal and Verguts (2011). Nevertheless, when we add the physical similarity variables, the influence of numerical distance disappears (see results of Experiment 1). This demonstrates the importance of including both numerical distance and physical similarity in such analyses.

In sum, our data likely differ from other published results because we (1) used many number stimuli, (2) did not categorize our numerical distances, and (3) assessed the influence of both numerical distance and physical similarity.

6.2. Automatic activation

In the present quantity same/different experiments, we assessed the activation of the quantity module under rigorous circumstances. Specifically, we assessed whether observers make the same/ different response across symbolic formats based on (a) the physical similarity of the two presented symbols or (b) the quantity denoted by the symbols. The results in this instance are clear: in all conditions, participants make the quantity same/different response across symbolic formats based on the physical similarity of the standard and the probe. There were no significant effects of numerical distance. This is the first time that such an effect has been demonstrated.

The lack of an influence of numerical distance on the cross-format quantity same/different tasks provides further evidence that numerals do not automatically activate their associated quantity representation (e.g., Cohen, 2009). The cross-format quantity same/different task is

a particularly rigorous assessment of automatic activation because on the surface there is no physical similarity between the numerals expressed in the two formats. The lack of a first order physical similarity favors a hypothesis that sameness is determined by comparing the numerals' quantities. So, despite there being a hypothesized necessity for the activation of the quantity module (let alone an irrelevant automatic activation for the task), there was no significant influence of numerical distance on the data. Thus, the data suggests that the numerals did not automatically activate the quantity representation module.

In addition to addressing the automatic activation of the quantity module, our data also addresses the activation of the symbolic format modules. We can address the automatic activation of the symbolic format modules with the same logic used to address the automatic activation of the quantity module. Here, we use the magnitude comparison task to assess the activation of the symbolic format modules. The data from the cross-format quantity same/different task has demonstrated that direct links exist between the Arabic digit and written number word symbolic format modules and between the Arabic digit and auditory number word symbolic format modules. We assessed whether these symbolic format modules are activated in a magnitude comparison task.

As in the quantity same/different task, the cross-format conditions are the critical conditions. These conditions are critical because any effect of physical similarity in the same-format conditions may be explained away as a perceptual interference effect. For example, in the simultaneous condition of the digit-digit magnitude comparison task there is an effect of physical similarity. However, because both the standard and probe are visible and in Arabic digit form, one can claim that the PS effect results from the visual processing of the Arabic digit standard interfering with the visual processing of the Arabic digit probe. Such a conclusion, which the data cannot refute, may not speak to the automatic activation of the symbolic format module. Such processing interference cannot manifest in the cross-format or successive conditions (in the digit-digit successive presentation condition, there remained an influence of physical similarity), therefore a PS effect would indicate that the symbolic format module was activated.

In all cross-format conditions of the magnitude comparison task except one, there was a significant effect of physical similarity. This indicates that the symbolic modules were activated in the cross-format magnitude comparison task. Here, one can conclude either (a) the symbolic modules were automatically activated and had no role in determining relative quantity or (b) the symbolic modules were activated because they were actively involved in processing the information required to determine relative quantity. The present set of experiments cannot adjudicate between these two conclusions. Nevertheless, there was no evidence for automatic activation of modules in the other cross-format conditions in which automatic activation was possible or even likely (e.g., the cross-format quantity same/different task). Therefore, we suggest that automatic activation is unlikely in this case as well. We hypothesize that the symbolic modules are activated here because they are involved in processing the information required to determine relative quantity. We explain this processing below in Section 6.3.

Finally, there is some empirical evidence to suggest that the quantity representations that are automatically activated are different than those that are intentionally activated (Bugden & Ansari, 2011; Tzelgov, Henik, & Berger, 1992; Tzelgov, Meyer, & Henik, 1992). Specifically, the automatically activated quantity representations may provide simple, dichotomous "large-small" information. The present paradigm was not designed to detect the presence of such a dichotomous representation, therefore the present data cannot address this issue. Nevertheless, flexible quantity representations are not inconsistent with the Multiple Representation Model (e.g., Cohen Kadosh & Walsh, 2009). Indeed, recently

Warren and Cohen (2013) have provided empirical evidence for flexible continuous quantity representations. Future research should address this issue.

In sum, the present data suggest that there is no automatic activation of either the relative quantity module or the symbolic format modules. Rather, these modules are activated when they are actively involved in processing information required to complete the task. Below, we interpret these findings within the context of different architectures of numerical cognition.

6.3. Numerical cognition architectures

The present set of experiments provides some implications for the architecture of the numerical cognition system. These implications are relatively clear for the description of the relation between the symbolic format modules. The conclusions one draws with regard to whether there exist one or more quantity modules linked to these symbolic format modules, however, depends on the conclusions one drew concerning the automatic activation of modules in general. We discuss each of these issues below.

To assess the relation between the symbolic format modules, we conducted cross-format same/different tasks. If there exists a direct link between the symbolic format modules, as predicted by both the Triple Code Model and the Multiple Representation Model, then RT should be predicted by Physical Similarity but not the Numerical Distance. This is exactly what we found. Importantly, in the cross-format conditions, there are no apparent physical similarities between the two stimuli. Nevertheless, the PS functions, which are based on the perceptual qualities of the different formats, predicted participants' RTs in all cross-format conditions. For the system to complete the task in this way, the auditory number words, written number words, and Arabic digits would have to directly activate the perceptual representations of each of the other forms. Thus, the data reveal direct links between the symbolic modules. This architectural feature is consistent with both the Triple Code Model and the Multiple Representation Model.

The Triple Code Model and the Multiple Representation Model each make different predictions for the magnitude comparison tasks. Specifically, the Triple Code Model proposes a single quantity representation that is shared by all format modules. Therefore, to complete the cross-format magnitude comparison task, this abstract representation will be activated and RT should be predicted solely by Numerical Distance. In contrast, the Multiple Representation Model proposes that each format module has a unique quantity representation. Therefore, to complete the cross-format magnitude comparison task, the two numerical symbols must first be converted into the same format before magnitude comparison can occur. Because both the symbolic format module and the quantity module will be activated, RT should be predicted by both Physical Similarity and Numerical Distance.

The description of the numerical cognition architecture that one can glean from the magnitude comparison tasks is in some ways dependent on the conclusions one draws from the data concerning the automatic activation of modules. In the current experiments, there was a significant effect of both physical similarity and numerical distance in five of the six relative quantity conditions. The one condition in which no physical similarity effect was found (i.e., the digit–word condition) could be interpreted as the single “tell-tale” result that unequivocally describes the numerical cognition architecture or the “anomaly” that simply reveals imperfect experimental conditions, statistics, etc. If one accepts the digit–word magnitude comparison result as a “tell-tale,” then the data indicate that the digit and word conditions share a quantity representation. That is, although the quantity same/different task revealed that the written number word and Arabic digit symbolic modules were directly

linked, that link was not activated in this condition of the magnitude comparison task. If the two symbolic formats did not share a quantity representation, then one would expect that the standard auditory symbol would activate its associated Arabic digit representation so a within format quantity comparison could occur. Because this did not happen, one may assume that both formats shared a common representation.

In contrast, if one accepts the digit–word magnitude comparison result as an “anomaly,” (as we do here) then one must make an assumption about automaticity before one can address the question of shared quantity representations. If one accepts that the cross-format activation is automatic, then the presence of the PS function is relatively meaningless with regards to the architecture of the numerical cognition system. There is, however, little evidence that the numerical cognition system automatically activates modules as an epiphenomenon. Rather, the system activates modules when these modules are involved in processing the information required to complete the task. Here, the system efficiently completes the cross-format magnitude comparison task in a two-step process. First the standard and/ or the probe activate a shared perceptual representation (e.g., Arabic digit) of the number symbol. Second, the standard and the probe’s shared, perceptual representation (e.g., Arabic digit) activate their associated quantities in the shared quantity representation module for comparison. In this way, both the PS and Welford function are relevant for the completion of the task, presumably because the numerical symbols do not share a common quantity representation.

Finally, neuroanatomical research assessing the abstract representation hypothesis has found mixed results. For example, using an fMRI, Piazza, Izard, Pinel, Bihan, and Dehaene (2004) demonstrated that different symbolic formats activate an abstract quantity representation in the hIPS. In addition, fMRI research has also found that multiple numerical formats activate the same IPS area (Jacob & Nieder, 2009; Piazza, Pinel, Bihan, & Dehaene, 2007). In contrast, fMRI research has found that quantity representations appear to be format dependent in both the left (Cohen Kadosh, Muggleton, Silvanto, & Walsh, 2010; Cohen Kadosh et al., 2007) and right (Cohen Kadosh et al., 2010) parietal lobes. Given the limits of fMRI research (e.g., Henson, 2005), functional data is critical to the interpretation of these data. Our results provide such functional data and support the existence of format dependent quantity representations.

6.4. Other implications

The present results reveal a clear effect of cross-format physical similarity. Thus, even without a first order physical similarity, physical similarity effects can arise because symbols can activate other representations. This finding is important to the numerical cognition research for practical reasons. That is, researchers may be tempted to use cross-format stimuli in an effort to inhibit physical similarity effects. Our data reveal that such a strategy, while on the surface is sound, in practice is invalid. Thus, we offer a cautionary note to researchers to assess physical similarity effects even in cross-format experiments.

As we show here, it is exceedingly difficult to distinguish between the “multiple representation with cross-format symbolic module activation hypothesis” and the “shared representation hypothesis.” We demonstrate that cross-format symbolic module activation is both viable and efficient. Because this cross-format activation may occur in a variety of untested experimental situations, researchers must consider this possibility before drawing conclusions concerning shared quantity representations. Cohen Kadosh and Walsh (2009) present a detailed review of the evidence in favor of a non-abstract quantity representation and how researchers may have misinterpreted evidence as favoring an abstract quantity representation.

6.5. Limitations

Here we assumed that numerical distance effect is an indicator of the activation of the quantity module and that the physical similarity effect is an indicator of the activation of the Arabic digit symbolic module. We also assumed that all modules necessary to complete the task would be activated by the task. Thus, we assumed modules that were not activated during a task were not necessary to complete the task. In general, these assumptions are readily accepted in the study of numerical cognition. Nevertheless, if our assumptions are invalid, our conclusions may be invalid as well. It should be noted that, after viewing the data, a pattern appeared that suggested that all activated modules were involved in completion of the task.

Furthermore, consistent with most models, it is possible to generate *a posteriori* alternative hypotheses to accommodate the current data and retain the “automatic activation” hypothesis. For example, it may be that form based decisions are faster than the quantity based decisions (and they run in parallel). Such alternative hypotheses should be tested in future work.

Another limitation is that, at this point, we cannot know in advance which surface representation will be activated in any particular task. The data suggest that the most likely (but not exclusively) surface representation to be activated is that of the probe. This may be because the observer is likely to convert the stimulus he or she attends to first into the format of the stimulus he or she attends to second. Both the position and predictability of the value of the stimulus can influence this. Thus when the standard is presented (a) in a privileged position (e.g., first or on top) and (b) is more predictable than the probe, then the probability the viewer will attend to the standard first is increased, thus increasing the probability the viewer will convert the standard into the probe. This, of course, is not the whole story because the data also show that observers will convert to an auditory representation, even when only visual stimuli are presented. More research is required to sort this issue out.

6.6. Conclusion

In sum, we conducted a quantity same/different task and quantity magnitude comparison task to explore how written number words, auditory number words, and Arabic digits are linked. In all conditions in the quantity same/different task, physical similarity was the primary controlling factor, while numerical distance had no influence. These data indicate there are direct, non-quantitative links between written number words, Arabic digits, and auditory number words and these links are preferred routes to determining the sameness of numerical symbols. Importantly, these effects were present in both same-format and cross-format conditions. Furthermore, our data indicate that both physical similarity and numerical distance are significant predictors in a traditional magnitude comparison task. The influence of physical similarity in the magnitude comparison task supports the conclusion that quantity representations are non-abstract.

Acknowledgments

This work supported by NIH Grant RO1HD047796 awarded to Dale J. Cohen.

References

- Banks WP, Hill DK. The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology Monograph*. 1974; 102:353–376.
- Buckley P, Gillman C. Comparisons of digits and dot patterns. *Journal of Experimental Psychology*. 1974; 103(6):1131–1136. [PubMed: 4457588]

- Bugden S, Ansari D. Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*. 2011; 118:32–44. [PubMed: 20970782]
- Campbell JID, Clark JM. An encoding complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*. 1988; 117(2):204–214.
- Chen Q, Verguts T. Beyond the mental number line: A neural network of number-space interactions. *Cognitive Psychology*. 2010; 60:218–240. [PubMed: 20138261]
- Cohen DJ. Integers do not automatically activate their quantity representation. *Psychonomic Bulletin and Review*. 2009; 16(2):332–336. [PubMed: 19293103]
- Cohen DJ. Evidence for direct retrieval of relative quantity information in a quantity judgment task: Decimals, integers, and the role of physical similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36(6):1389–1398.
- Cohen DJ, Ferrell JM, Johnson N. What very small numbers mean. *Journal of Experimental Psychology: General*. 2002; 131(3):424–442. [PubMed: 12214756]
- Cohen Kadosh R, Cohen Kadosh K, Kaas A, Henik A, Goebel R. Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron*. 2007; 53:307–314. [PubMed: 17224410]
- Cohen Kadosh R, Muggleton N, Silvanto J, Walsh V. Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral Cortex*. 2010; 20(9):2166–2171. [PubMed: 20051361]
- Cohen Kadosh R, Walsh V. Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*. 2009; 32:313–373. [PubMed: 19712504]
- Cohen L, Dhaene S, Chochon F, Lehericy S, Naccache L. Language calculation within the parietal lobe: A combined cognitive, anatomical, and fMRI study. *Neuropsychologia*. 2000; 38:1426–1440.
- Dehaene S. The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception and Psychophysics*. 1989; 45:557–566. [PubMed: 2740196]
- Dehaene S. Varieties of numerical abilities. *Cognition*. 1992; 44(1–2):1–42. [PubMed: 1511583]
- Dehaene, S. *The number sense: How the mind creates mathematics*. New York: Oxford University Press; 1997.
- Dehaene S, Akhavein R. Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21(2):314–326.
- Dehaene S, Bossini S, Giraux P. The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*. 1993; 122:371–396.
- Dehaene S, Dupoux E, Mehler J. Is numerical comparison digital? Analogical and symbolic effects in two-digit comparison. *Journal of Experimental Psychology: Human Perception and Performance*. 1990; 16:626–641. <http://dx.doi.org/10.1037/0096-1523.16.3.626>. [PubMed: 2144576]
- Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*. 1999; 284:970–974. [PubMed: 10320379]
- Dehaene S, Tzourio N, Frak V, Raynaud L, Cohen L, Mehler J, et al. Cerebral activations during number multiplication and comparison: A PET study. *Neuropsychologia*. 1996; 34:1097–1106.
- Duncan EM, McFarland CE. Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory and Cognition*. 1980; 8:612–622. [PubMed: 6163942]
- Gallistel CR, Gelman R. Preverbal and verbal counting and computation. *Cognition*. 1992; 44(1):43–74. [PubMed: 1511586]
- Gallistel CR, Gelman R. Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*. 2000; 4:59–65. [PubMed: 10652523]
- Ganor-Stern D, Tzelgov J. Across-notation automatic numerical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34:430–437.

- Ganor-Stern D, Tzelgov J, Ellenbogen R. Automaticity and two-digit numbers. *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33(2):483–496. <http://dx.doi.org/10.1037/0096-1523.33.2.483>. [PubMed: 17469981]
- Garcfa-Orza J, Perea M, Mallouh RA, Carreiras M. Physical similarity (and not quantity representation) drives perceptual comparison of numbers: Evidence from two Indian notations. *Psychonomic Bulletin and Review*. 2012; 19:294–300. [PubMed: 22271341]
- Geyer LH. Recognition and confusion of the lowercase alphabet. *Perception and Psychophysics*. 1977; 22(5):487–490.
- Gibbon J, Church RM. Time left: Linear versus logarithmic subjective time. *Journal of the Experimental Analysis of Behavior*. 1981; 7:87–107.
- Goldfarb L, Henik A, Rubinsten O, Bloch-David Y, Gertner L. The numerical distance effect is task dependent. *Memory and Cognition*. 2011; 39:1508–1517. [PubMed: 21698497]
- Gonzalez EG, Kollers PA. Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1982; 8(4):308–319.
- Henson R. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*. 2005; 58:193–233.
- Hinrichs J, Yurko D, Hu J. Two digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7(4):890–901.
- Jacob S, Nieder A. Notation-independent representation of fractions in the human parietal cortex. *The Journal of Neuroscience*. 2009; 29(14):4652–4657. [PubMed: 19357289]
- McCloskey M. Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*. 1992; 44:107–157. [PubMed: 1511584]
- McCloskey M, Caramazza A, Basili A. Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*. 1985; 4:171–196. [PubMed: 2409994]
- McCloskey M, Macaruso P. Representing and using numerical information. *American Psychologist*. 1995; 50:351–363. [PubMed: 7762888]
- McCloskey M, Sokol S, Goodman R. Cognitive processes in verbal-number production: Inferences from the performance of brain damaged subjects. *Journal of Experimental Psychology: General*. 1986; 115(4):307–330. [PubMed: 2949043]
- McCusker LX, Gough PB, Bias RG. Word recognition inside out and outside in. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7:538–551.
- Moyer R, Landauer T. Time required for judgments of numerical inequality. *Nature*. 1967; 215(5109):1519–1520. [PubMed: 6052760]
- Moyer RS, Landauer TK. Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*. 1973; 1:167–168.
- Nieder A, Miller EK. Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*. 2003; 37:149–157. [PubMed: 12526780]
- Piazza M, Izard V, Pinel P, Bihan D, Dehaene S. Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*. 2004; 44:547–555. [PubMed: 15504333]
- Piazza M, Pinel P, Bihan D, Dehaene S. A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*. 2007; 53:293–305. [PubMed: 17224409]
- Price GR, Ansari D. Symbol processing in the left angular gyrus: Evidence from passive perception of digits. *Neuroimage*. 2011; 57(3):1205–1211. [PubMed: 21620978]
- Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. *Psychological Review*. 1989; 96(4):523–568. [PubMed: 2798649]
- Shepard RN, Kilpatrick DW, Cunningham JP. The internal representation of numbers. *Cognitive Psychology*. 1975; 7:82–138. [http://dx.doi.org/10.1016/0010-0285\(75\)90006-7](http://dx.doi.org/10.1016/0010-0285(75)90006-7).
- Takahashi A, Green D. Numerical judgments with Kanji and Kana. *Neuropsychologia*. 1983; 21(3):259–263. [PubMed: 6877579]
- Tzelgov J.; Ganor-Stern, D. Automaticity in processing ordinal information. In: Campbell, JID., editor. *Handbook of mathematical cognition*. New York: Psychology Press; 2005. p. 55-67.
- Tzelgov J, Henik A, Berger J. Controlling stroop effects by manipulating expectations for color words. *Memory and Cognition*. 1992; 20:727–735. [PubMed: 1435275]

- Tzelgov J, Meyer J, Henik A. Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1992; 18:166–179.
- Van Opstal F, Verguts T. The origins of the numerical distance effect: The same-different task. *Journal of Cognitive Psychology*. 2011; 23:112–120.
- Verguts T, Van Opstal F. Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin and Review*. 2005; 12:925–930. [PubMed: 16524012]
- Warren E, Cohen DJ. On the relativity of relative frequencies. *Attention, Perception, and Psychophysics*. 2013; 75:614–629.
- Welford AT. The measurement of sensory-motor performance: Survey and reappraisals of twelve years' progress. *Ergonomics*. 1960; 3:189–230.

Fig. 1.

Three theoretical architectures of numerical cognition. The architecture at the top of the figure depicts a simplified version of McCloskey's single representation model. The architecture at the center of the figure depicts a simplified version of Dehaene's Triple Code Model. The architecture at the bottom of the figure depicts a strict multiple representation architecture. The rounded rectangles represent modules that process numeric symbols in specific formats. The circles represent modules that process quantity information (QR). The arrows represent direct links between the processing modules. See text for detailed descriptions of each architecture.

Fig. 2.

An illustration of the physical similarity of the number words “two” and five.”

Fig. 3.

Boxplots of the slopes of the predictor variables for the digit–digit, word–digit, digit–word, and word–word conditions of Experiment 1. The horizontal grey line indicates a slope of zero (i.e., no effect). The boxplots describe the entire distribution of slopes from the participants for each predictor. The box denotes the range of the middle 50% of the data, the internal line denotes the position of the median, and the whiskers extend out to the tail of the distribution. The circles indicate points that fall outside the densest cluster of data. The boxplots reveal that, in each condition, a physical similarity predictor robustly predicts performance while the remaining predictors remain centered on zero.

Fig. 4. Boxplots of the slopes of the predictor variables for the digit-digit, word-digit, digit-word, and word-word conditions of Experiment 2. The horizontal grey line indicates a slope of zero (i.e., no effect).

Fig. 5. Boxplots of the slopes of the predictor variables for the digit–digit and auditory–digit conditions of Experiment 3. The horizontal grey line indicates a slope of zero (i.e., no effect).

Fig. 6. Boxplots of the slopes of the predictor variables for the digit–digit and auditory–digit conditions of Experiment 4. The horizontal grey line indicates a slope of zero (i.e., no effect).

Table 1

The physical similarity functions between two written number words (PS_{word}), auditory number words (PS_{auditory}), and two Arabic digits (PS_{digit}), for the standards and probes used in Experiments 1–4 (rounded to three digits).

Standard	Probe	PS_{word}	PS_{auditory}	PS_{digit}
2	1	7.028	6	0.2
2	3	14.933	7	2
2	4	3.563	5	0.4
2	5	11.333	3	0.75
2	6	3.194	1	1.333
2	7	4.15	3	0.5
2	8	1.4	1	2.5
2	9	8.813	4	0.75
5	1	8.917	3	0.2
5	2	11.333	3	0.75
5	3	7.967	5	2
5	4	10.641	7	1
5	6	6.771	3	5
5	7	5.0417	3	0.5
5	8	3.746	7	2.5
5	9	10.907	5	2
8	1	8.567	3	0.4
8	2	1.4	3	2.5
8	3	5.696	3	2.5
8	4	4.796	4	1.333
8	5	3.746	7	2.5
8	6	4.667	5	6
8	7	3.291	3	0.75
8	9	4.088	7	2.5

Table 2

The correlations (and *t*-tests) between the physical similarity measures and the Welford function for the standards 2, 5, and 8, and the probes 1–9.

	PS_{word}	PS_{auditory}	PS_{digit}
<i>Welford</i>			
	–0.05	0.46	0.40
	$t(22) = 0.23, ns$	$t(22) = 2.45, p = .02$	$t(22) = 1.99, p = .06$
<i>PS_{digit}</i>			
	–0.22	0.14	
	$t(22) = 1.05, ns$	$t(22) = 0.64, ns$	
<i>PS_{auditory}</i>			
	0.26		
	$t(22) = 1.27, ns$		

Table 3

Predictions of each major model of numerical architecture for the primary predictor in the cross-format quantity same/different and magnitude comparison tasks.

Model	Task	<u>Significant predictor</u>	
		Physical similarity	Numerical distance
McCloskey single representation model	Same/different	No	Yes
	Magnitude comparison	No	Yes
Dehaene Triple Code Model	Same/different	Yes	No
	Magnitude comparison	No	Yes
Multiple Representation Model	Same/different	Yes	No
	Magnitude comparison	Yes	Yes

Table 4

The means (sds) and error rates for each condition of Experiments 1–4.

Condition	Task	RT	Error
<i>Cross-format conditions</i>			
Word–digit (Experiments 1 and 2)	Same/different	834 (335)	0.04
	Magnitude comparison	1014(555)	0.06
Digit–word (Experiments 1 and 2)	Same/different	839 (318)	0.03
	Magnitude comparison	996 (544)	0.06
Auditory–digit (Experiments 3 and 4)	Same/different	817 (446)	0.03
	Magnitude comparison	830(481)	0.06
<i>Same-format conditions</i>			
Digit–digit (Experiments 1 and 2)	Same/different	683 (269)	0.04
	Magnitude comparison	809 (455)	0.04
Word–word (Experiments 1 and 2)	Same/different	727 (302)	0.04
	Magnitude comparison	1070(555)	0.06
Digit–digit (Experiments 3 and 4)	Same/different	743 (412)	0.03
	Magnitude comparison	756 (446)	0.05

Table 5

The significant predictors for each condition of Experiments 1–4.

Condition	Task	Significant predictor	
		Physical similarity	Numerical distance
Cross-format conditions			
Word–digit (Experiments 1 and 2)	Same/different	Yes	No
	Magnitude comparison	Yes	Yes
Digit–word (Experiments 1 and 2)	Same/different	Yes	No
	Magnitude comparison	No	Yes
Auditory–digit (Experiments 3 and 4)	Same/different	Yes	No
	Magnitude comparison	Yes	Yes
Same-format conditions			
Digit–digit (Experiments 1 and 2)	Same/different	Yes	No
	Magnitude comparison	Yes	Yes
Word–word (Experiments 1 and 2)	Same/different	Yes	No
	Magnitude comparison	Yes	Yes
Digit–digit (Experiments 3 and 4)	Same/different	Yes	No
	Magnitude comparison	Trend	Yes

Table 6

The slopes and *t*-values for all predictors for all Experiments 1–4.

Predictor	Same/different Experiment 1		Magnitude comparison Experiment 2	
	Slope	<i>t</i>	Slope	<i>t</i>
<i>Digit–digit</i>				
Welford	0.01	0.4	0.1	8.7**
PS _{digit}	0.09	9.2**	0.04	6.1**
PS _{word}	0.0	−0.3	0.01	1.7
PS _{auditory}	0.05	4.2**	0.0	0.5
<i>Word–digit</i>				
Welford	0.01	0.09	0.12	13.1**
PS _{digit}	0.06	6.3**	0.05	4.1**
PS _{word}	−0.01	−1.1	0.03	3.5**
PS _{auditory}	0.02	1.2	0.01	0.5
<i>Word–word</i>				
Welford	0.0	0.1	0.1	8.7**
PS _{digit}	0.0	−0.4	−0.01	−0.7
PS _{word}	−0.01	−0.6	0.01	0.8
PS _{auditory}	0.05	4.3**	0.01	0.6
<i>Digit–word</i>				
Welford	0.3	2.2	0.14	12.5**
PS _{digit}	−0.2	−2.4	0.01	1.1
PS _{word}	0.7	7.5**	0.04	5.3**
PS _{auditory}	−0.1	−1.0	−0.02	−2.1
	Experiment 3		Experiment 3	
<i>Digit–digit</i>				
Welford	−0.01	−1.0	0.03	2.5*
PS _{digit}	0.5	4.9**	0.02	1.2
PS _{word}	0.0	0.3	0.03	2.4
PS _{auditory}	0.0	0.4	0.01	2.4
<i>Auditory–digit</i>				
Welford	−0.02	−2.2	0.07	4.8**
PS _{digit}	0.04	2.5*	0.04	2.4
PS _{word}	0.03	2.4	0.04	2.9*
PS _{auditory}	0.0	0.3	0.01	1.1

*
p < 0.05.

**
p < 0.01.