A look behind perceptual performance in numerical cognition

Thesis by

Dennis Reike

Submitted to the Human Sciences Faculty

for the Degree of

Dr. phil.

Submission: November 2017

University of Potsdam

Supervisor: Prof. Dr. rer. nat. Wolfgang Schwarz

Cosupervisor: Dr. habil. Eberhard Schröder

This is to my beloved children Timo and Lucy.

If you ever have the idea to read this you should know that the only reason to do this is that you are really interested in numerical cognition.

Publication overview

This thesis is based on following publications:

Reike, D., & Schwarz, W. (2016). One model fits all: Explaining many aspects of number comparison within a single coherent model – A random walk account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**, 1957-1971. DOI: 10.1037/xlm0000287

Reike, D., & Schwarz, W. (2017). Exploring the origin of the number size congruency effect: Sensitivity or response bias? *Attention, Perception, & Psychophysics*, **79**, 383-388. DOI: 10.3758/s13414-016-1267-4

Schwarz, W., & Reike, D. (2017). Local probability effects of repeating irrelevant attributes. *Attention, Perception, & Psychophysics*, **79**, 230-242. DOI: 10.3758/s13414-016-1200-x

Contents

Chapter 1

Introduction:

Specific aspects of number processing

One two three

A little fool I want to be

Two three four

You can give me more

Five six seven

I don't want to wait for heaven

Nine ten eleven

–

Going back to seven

Seven eight nine

Kann denn das noch sein?

Blank & Meier (Yellow)

Symbols represent the most amount of information in a compressed way. They are used in many different situations as powerful tools for communication and to solve a variety of problems. As humans, we know at least two different types of symbols. Many symbols consist of features that are directly accessible. They have a selfexplanatory character, like a pictogram of a crossed mobile. The meaning of other symbols has to be learned and without this learning process we are unable to understand these types of symbols. One example of such a symbol is a digit. Digits represent many features of quantities, and after learning these features, we connect them to the digits. Subsequently, we are able to easily understand and use digits as numbers.

Humans have a considerable ability to recognize and respond in relation to any kind of quantitative information. We encounter quantities in many different forms and with various intentions. For example, we communicate temperature or an amount of fluid as numerical values. Similarly, if we buy pieces of anything we have to count, we pay a price that reflects a quantity. It seems easy and uncomplicated to manage this large diversity of numerical information and we use the same symbols for all these different forms and usages – digits.

One approach to investigate number processing is to consider number comparison, which is an important aspect of it. It seems that comparing quantities is a fundamental cognitive skill of considerable survival value for animals and humans. However, all creatures have to decide several behavioral matters each day and many of these decisions are in principle based on quantity comparison. For example, for animals it is very important to know how many members belong to a potentially hostile pack in comparison to their own pack. When the foreign pack is smaller, maybe it is worth fighting; if it is larger, then it could be better to flee. This is a very drastic example to demonstrate the evolutionary advantage for creatures that have the ability to count and compare numbers.

In fact, groups of lions adjust their agnostic behaviour depending on the amount of their own number in comparison to the number of a foreign group. McComb, Packer, and Pusey (1994) reported that defending lionesses were more likely to approach roaring playbacks of one foreign lioness than of a group of three. In addition, they approach a group of three more cautiously. Furthermore, the smaller their own group, the more careful the group was when approaching the playback. If several members of the own group were absent, then they were tempted to recruit extra companions by roaring. They did so more likely the smaller the own and the larger the foreign group was.

Is this observed performance unique, or do other animals show comparable performance in comparing quantities? Indeed, it seems that several species have the ability to recognize quantities and use them for their decisions. For example, pigeons have the ability to compare two small quantities (Brannon, Wusthoff, Gallistel, & Gibbon, 2001), even if one of the two quantities remains after a numerical subtraction and should be compared to a constant quantity. Pigeons are also able to use symbols for quantities. Xia, Emmerton, Siemann, and Delius (2001) reported that trained pigeons responded to numerosities by choosing the related symbol. Another species that shows an ability to compare quantities, is monkeys. Nieder, Freedman, and Miller (2002) reported that monkeys realize and indicate whether two successively presented quantities of items contain the same quantity. Monkeys have the ability to detect ordinal disparity of numerosities (Brannon & Terrace, 1998), and they can be trained to use symbols to name the number of items in a display (Livingstone, Srihasam, & Morocz, 2010; Matsuzawa, 1985). Pigeons and monkeys are not the only animals that have these numerical skills. Parrots can label numerosities (Pepperberg, 1987). Chicks show human-like behavior in relation to numbers (Rugani, Vallortigara, Priftis, & Regolin, 2015). Finally, crows (Ditz, & Nieder, 2016), squirrels (Hassmann, 1952) and even fish (Agrillo, Dadda, Serena, & Bisazza, 2008) can compare quantities as well.
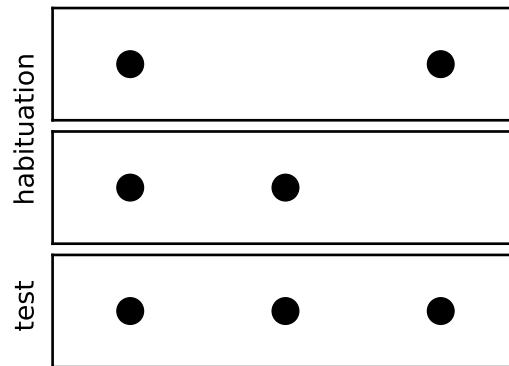
Figure 1. Example stimuli of dot comparison as used with infants. The upper two panels represent stimuli for the habituation period with two dots. The bottom panel shows a test stimulus with three dots. (Adopted from Starkey & Cooper, 1980)

Given that broad account of numerical cognition in animals, it is not surprising that similar results can be found for humans. Specifically, do human infants show reliable performance in numerical cognition too? Starkey and Cooper (1980) reported the ability of 6-month-old babies to distinguish small quantities. In the habituation period, they bored infants when presenting a display series with the same numbers of dots. They observed for how long the babies stared at these dots. Finally, a different number of dots were presented (see Figure 1). The babies stared longer on three dots as on two dots in the habituation period (for recent review and extensions see, e.g., Mou & vanMarle, 2014; Starkey & McCandliss, 2014). This was interpreted to show that the babies were surprised of the unexpected quantity, and, therefore, they inspected it longer. The babies could distinguish two and three, but not four and six dotes. The same result occurs when different stimuli materials were used such as some animals instead of dots (Strauss & Curtis, 1981). In addition, very young babies (in the first week of their lives) are able to discriminate between two and three dots (Antell & Keating, 1983). Furthermore, children in the kindergarten and in school show considerable ability to compare two numerosities or numbers (e.g., Holloway & Ansari, 2009; Mussolin, Mejias, & Noël, 2010; Sekuler & Mierkiewicz, 1977). Number comparison as a main topic of this thesis will be considered in more detail below.

In addition, Wynn (1992) reported that five-month-old infants have the ability to understand simple arithmetic procedures. To test the innate understanding of addition, babies had to observe how two dolls were separately placed behind a screen. Then, the screen was removed and the babies were confronted with one of two situations. Meanwhile, the time was measured to see how long the infants watched the yielded dolls. In one condition, the expected result of two dolls was presented; in the other condition, only one doll was presented. To test the understanding of subtraction, a similar procedure was used. Two dolls were seen in the beginning, and then they were covered by a screen. The babies observed the removing of one doll. The infants stared longer at one doll in the addition task and at two dolls in the subtraction task. This was interpreted to mean that the infants were surprised (or irritated) in these conditions and therefore have a basic understanding of simple arithmetic even at very young ages.

What are the numerical and mathematical skills of children in school? Children show different skills from the first grade on, for example digit naming, number comparison, dot enumeration, number line estimation, addition and subtraction (e.g., Moore & Ashcraft, 2015). Over schooldays, children continue to develop these skills. However, children show considerable performance in numerical tasks. To consider adults' numerical cognition, different aspects (and effects) will be presented separately.

Number Comparison

Numerical distance effect. Moyer and Landauer (1967) reported first that the time required to decide which of two presented digits is numerically larger decreases with increasing numerical distance of these two digits, the numerical distance effect. In addition, the error rate decreases with increasing numerical distance. Although the response time is faster in digit comparison than in comparing

numerosities, the basic pattern (including the numerical distance effect) in digit comparison is similar to comparing numerosities (e.g., Buckley & Gillman, 1974; Ratcliff, Thompson, & McKoon, 2015). The numerical distance effect was demonstrated in different notations, for example in Japanese scripts Kanji and Kana (Takahashi & Green, 1983) or in Hebrew (Razpurker-Apfeld & Koriat, 2006).

Is this effect limited to a strict number comparison tasks? Schwarz and Eiselt (2012) reported a numerical distance effect also in a visual search task. The participants had to indicate whether there was a 5 within presented digits. The targetdistractors were presented in a condition numerically close (3,4,6,7) to the 5 or far (1,2,8,9). Participants solved the task faster and were less error-prone in the numerically far condition.

Magnitude effect. Very similar to the numerical distance effect is the magnitude effect. This is the time required to indicate the numerically larger number out of two increases with increasing numerical magnitude by constant numerical distances (e.g., Buckley & Gillman, 1974; Moyer & Landauer, 1967). In addition, the error rate decreases with increasing numerical magnitude.

The magnitude effect leads to the strong assumption that perceiving numbers is similar to perceiving physical properties due to the Weber–Fechner law. To decide whether two objects differ in a physical property, a minimum difference is needed between these two objects related to this property. Furthermore, this minimal difference for discrimination depends on the actual value of the physical property of both objects. Specifically, the larger one of the actual values of the physical property is the larger the difference between both objects has to be. For example, consider a minimum difference in length to distinguish two given lines. Next, if the smaller line length is doubled, then the larger line length (and thereby the minimum difference) has to be doubled also to be able to decide that these longer lines have different lengths rather than being similar (Festinger, 1943).

To consider this in relation to the perception of numbers, it is useful to take into account their mental representation. Can the manner of the mental representation of numbers provide us with an explanation for the magnitude effect and for the numerical distance effect? To answer this question, it is necessary to consider what we already know about mental number representation.

The mental number line

Spatial-numerical association. At school, children learn a number line. Usually, numbers are sorted from smallest to largest left to right on this line. Dehaene, Bossini, and Giraux (1993) let participants indicate by pressing a button whether numbers were odd or even. Participants solved this task with the left hand faster for small digits than for large digits and with the right hand faster for large digits than for small digits, which was termed the spatial-numerical association of response code (SNARC) effect. This was often seen as evidence for an internalized representation of the number line – a mental number line. The usage of this mental number line seems flexible and adaptable to the presented range (Dehaene et al., 1993). Within a range from 0 to 5, the 5 is located on the right side and within a range from 4 to 9 the 5 is located on the left side. In addition, the SNARC effect can be extended to multi-digit numbers (Nuerk, Moeller, Klein, Willmes, & Fischer, 2011) and does not depend on specific effectors (Schwarz & Keus, 2004; Schwarz & Müller, 2006). Furthermore, the SNARC effect occurs also in other tasks. For example, in a simple detection experiment, the spatial attention shifts to the left after small and to the right after large numbers (Fischer, Castel, Dodd, & Pratt, 2003).

Linear vs. logarithmic. The specific nature of the mental representation of numbers is broadly discussed in the literature. Specially, for the scale of the mental number line, mostly two types of relations from numbers to the mental represen-

tation were suggested – a linear and a logarithmic number line, as illustrated in Figure 2.
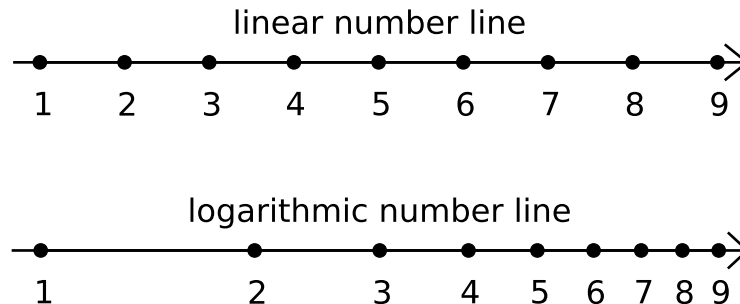
linear number line

1 2 3 4 5 6 7 8 9

logarithmic number line

1   2   3   4 5 6 7 8 9

Figure 2. Illustration of linear and logarithmic number lines. The upper line shows a linear relation of numbers whereas the lower line shows a logarithmic relation. The linear number line demonstrates that the perceived difference between two neighboring numbers is equal for all pairs. In the logarithmic number line the perceived difference between two neighboring numbers decreases with increasing magnitude.

In a number naming task, Whalen, Gallistel, and Gelman (1999) let participants press as many key presses as the magnitudes value presented. Humans showed the same result as animals. The variance of the answers increases with increasing magnitude. This supports the view of a logarithmic internal representation of the mental number line. Therefore, the perception of numbers is likely to be comparable to the perception of physical features, due to the Weber–Fechner Law. Huber, Moeller, and Nuerk (2014) reported that the scale of the mental number line seems to be flexible for adults who adapted it to the specific task. Children did not show this flexibility.

Is the numerical relation between numbers and space the only relation concerning the perception of magnitudes? Bonato, Zorzi and Umiltà (2012) presented empirical evidence also for a relation between space and time – a mental time line. However, this suggests that there may be a more complex magnitude representation system.

A theory of magnitude

Walsh (2003) proposed a model that integrates different aspects and variants of magnitude. He suggested shared cognitive processes to perceive space, time and quantity. Henik, Leibovich, Naparstek, Diesendruck, and Rubinsten (2012) suggested a direct relation from this analogous representation of magnitude to the numerical core system. For example, the SNARC effect supports the view that there is a cognitive inference between space and numbers. Are there inferences between other dimensions, specifically between different quantity features?

The number-size congruency effect. In a study by Besner and Coltheart (1979), participants had to decide which of two presented digits with different physical sizes was numerically larger. The participants were faster and less afflicted with errors in conditions with congruent numerical magnitude and physical sizes (e.g., 2 − 8) than in incongruent conditions (e.g., 2 − 8), the number-size congruency effect. Henik and Tzelgov (1982) reported the same effect for the reversed task of deciding which of two presented digits was physically larger. The number-size congruency effect occurs also in other notation systems (Razpurker-Apfeld & Koriat, 2006) and different tasks like visual search (Sobel, Puri, & Faulkenberry, 2016). Presently, it is unclear which processing stage is involved in the cognitive process of deciding whether the physical size or the magnitude is responsible for the interaction of physical size and the magnitude in response performance. Schwarz and Heinze (1998) reported in an event-related potential study that lateralized readiness potentials are shifted in time depending on the inference of numerical and physical size. In a number-size congruity task participants showed a stronger activation in their dorsolateral prefrontal cortex and the anterior cingulate cortex in incongruent trials than in congruent trials (Kaufmann et al., 2005). What else do we know about the brain in relation to the number processing and representation beyond the number-size congruity effect?
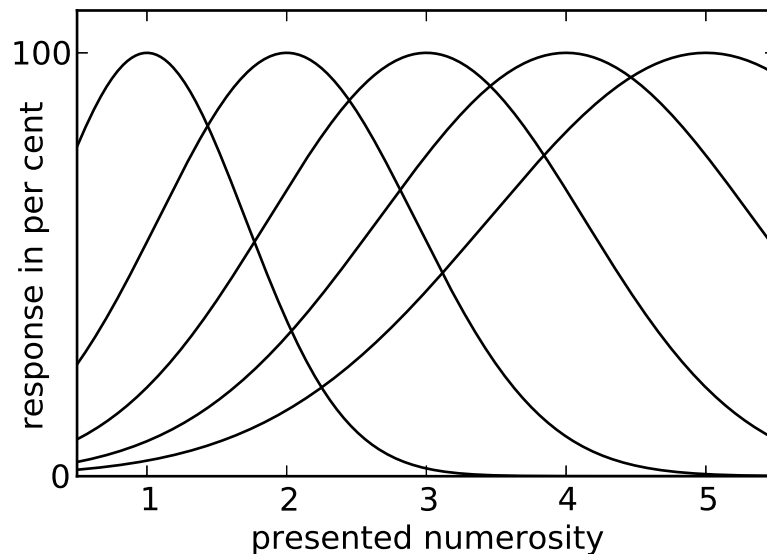
Numbers and the brain

Figure 3. Illustration of response rates of numerosity specific neurons. Approximated normalized response rates of neurons, which are specific to 1, 2, 3, 4 or 5 (from left to right) as a function of the presented numerosity. Every neuron reaches the 100 per cent response rate if the specific numerosity is presented. The response rate is declining the further the presented numerosity is to the specific numerosity. (Adopted from Nieder, 2002)

Dehaene, Piazza, Pinel, and Cohen (2003) proposed three parietal circuits for number processing. The horizontal segment of intraparietal sulcus is needed for quantity processing, the angular gyrus is needed for verbal processes in numeric relations and the posterior superior parietal lobule regulates spacial and non-spacial attention. In addition, a meta analysis from Arsalidou and Taylor (2011) suggests that in more complex tasks several different brain areas are involved, for instance frontal and prefrontal regions for arithmetic processing. However, they also reported that the inferior and superior parietal lobules are needed for numerical tasks. Kaufmann et al. (2005) showed a stronger activation in bilateral parietal areas, including the intraparietal sulcus, for smaller numerical distances (see also Dehaene, 2006). Furthermore, the parietal cortex is involved in number processing while solving different number tasks (Piazza & Eger, 2016). The inferior parietal

sulcus is used even for the processing of numerical symbols as digits (Holloway, Battista, Vogel, & Ansari, 2013).

Nieder (2002) demonstrated the existence of numerosity specific neurons. The response rate of selective neurons depends on the quantity presented. It is the strongest response rate at the preferred quantity and the larger numerical distance from this preferred quantity is the lower the response rate. As illustrated in Figure 3, the response rate of number sensitive neurons will be more imprecise with increasing numerical magnitude.

Number representation

In light of the above, three main conclusions will be suggested about number representation. First, numbers are represented in a noisy manner and these representations are sorted according to numeric value. Probably, the representations are noisier with increasing numerical magnitude, as illustrated in Figure 3. Therefore, number perception is analogous to the perception of physical properties due to the Weber–Fechner Law. Second, the number processing system shares parts with the general processing of magnitude. Third, we use strategic spatial information to understand numbers. Specifically, we sort numbers from left to right in increasing order. The spatial location probably helps to distinguish smaller from larger numbers.

Following these main conclusions, this thesis elaborates on three main questions. First, what kind of mental process is the number comparison process to produce effects reported above (e.g., the numerical distance effect) with the use of noisy representations of numbers? Second, does the interaction of numerical magnitude and physical size affect the actual perception, or does it affect the response part of the cognitive process? Third, can we use the mental representation of numbers and related effects (e.g., the SNARC effect) as tools to investigate other

relations? This thesis will answer these questions in more detail in three chapters, but they are summarized below.

Explaining many aspects of number comparison within one coherent model

Random walk and diffusion models are broadly discussed as process models in relation to decision tasks. Specifically, in two choice tasks where participants have to give one of two possible answers, random walk and diffusion models have been increasingly applied. For example, Ratcliff and Smith (2004) successfully fitted diffusion models to a signal detection task, a lexical decision task, and a recognition memory task. Ratcliff, Smith, Brown, and McKoon (2016) reviewed the broad application possibilities of random walk and diffusions models. In addition, activities of neurons in the superior colliculus were described within the framework of diffusion models (Ratcliff, Cherian, & Segraves, 2003). These fire rates were linked to the response times of eye movements of monkeys in a distance related two choice tasks. Gold and Shadlen (2007) trained monkeys to perform a motion detection task, which was also a two choice task, with eye movement as response. They explained neuronal activity (e.g., in the lateral intraparietal area) within the framework of random walk models.

Poltrock (1989) first used a random walk model as a framework to explain response times and error rates within a digit comparison task. He explained basic findings in numerical cognition like the numerical distance effect. Smith and Mewhort (1998) validated the usability of random walk models in number comparison tasks and extend further explanations as speed accuracy effects within the framework of random walk models. Schwarz and Ischebeck (2003) used a diffusion model that coalesced numerical magnitude and physical size. They described re-

sponse times and error rates quantitatively for a numerical and a physical size task within the framework of their coalescence model.

Random walk models provide a coherent chronometric framework, based on a dynamic, process-orientated view. In the second chapter of this thesis, a random walk model with parsimonious usage of meaningful interpretable parameters is used to account for open questions in relation to numerical comparison and the usability of random walk models for numerical comparison tasks.

It will be shown that random walk models account quantitatively for the full matrix of response times and error rates in a complete paired digit comparison design. Response time variance is quantitatively explained even for the full matrix of digit pairs. Furthermore, error response times are investigated in detail. The mean error response time of a given digit pair (e.g., 2–8) is faster than the mean correct response time of the complementary digit pair (e.g., 8–2) and random walk models can account for it. Specifically, different from standard assumptions often made in random walk models, this account requires that the distributions of step sizes of the induced random walks are asymmetric. Random walk models predict a numerical distance effect even for error response times and this effect clearly occurs in the observed data. Furthermore, the presented model provides a well-defined framework to investigate the nature and scale (e.g., linear vs. logarithmic) of the mapping of numerical magnitude onto its internal representation. In comparison of the fits of proposed models with linear and logarithmic mapping, the logarithmic mapping is suggested to be prioritized.

Finally, a novel oculomotor effect is reported, namely the saccadic overschoot effect. The participants responded by saccadic eye movements and the amplitude of these saccadic responses decreases with numerical distance. However, it will be discussed how the model used can help to interpret complex findings (e.g., conflicting speed vs. accuracy trends) in applied studies that use number comparison as a well-established diagnostic tool.

Exploring the origin of
the number-size congruency effect

As discussed above, the number-size congruency effect describes an interaction between numerical and physical size in tasks that involve numerical or physical size comparisons. Specifically, participants can solve such tasks faster in congruent conditions (e.g., $2 - 8$) than in incongruent conditions (e.g., $2 - 8$). An as-yet-open question is, whether the benefit in congruent conditions is related to a better perception than in incongruent conditions? Alternatively, the number-size congruency effect mediated response biases due to number or respectively physical size.

The signal detection theory is a perfect tool to distinguish between these two alternatives. It describes two parameters, namely sensitivity and response bias (Macmillan & Creelman, 2005). Changes in the sensitivity relate to the actual task performance due to real differences in perception processes whereas changes in the response bias simply reflect strategic implications as higher preparation of an anticipated answer.

In Chapter 3, the signal detection theory is applied to a task that required participants to judge the physical size of digits. The results clearly demonstrate that the number-size congruency effect cannot be reduced to mere response bias effects and that genuine sensitivity gains for congruent number-size pairings than contributes to the number-size congruency effect.

Local probability effects of
repeating irrelevant attributes

Stimuli with different attributes are often used in research experiments. These attributes, while logically and statistically independent, often produce mental conflicts within these attributes or in relation to the performed answers. For example,

Simon (1969) reported that responses in the direction of an irrelevant location performed faster than responses to the opposite site of the location, known as the Simon-effect. This effect directly relates to the effect that responses are faster given to smaller (larger) numbers on the left (right) hand than on the right (left) hand (the SNARC effect; Dehaene, Bossini, & Giraux, 1993).

Bertelson (1961) showed that the trial-by-trial repetition probability of a stimulus could be varied without altering the global (i.e., overall) probability of the stimulus to occur. In a two choice design he varied the response repetition independently from the global probability of stimuli occurrence. Responses in actual response repetitions showed a gain in performance in relation to response changes in high repetition conditions as compared to low repetition conditions.

Chapter 4 presents a research design that was not used with conflict tasks before. In a Simon and a SNARC task, the local transition probability of irrelevant attributes (location, magnitude) varied while local transition probability of relevant attributes (color, parity) and the global probability occurrence of each stimuli were kept constant. Participants are quite sensitive and able to recognize the underlying local transition probability of irrelevant attributes. They show a performance gain for actual repetitions of the irrelevant attribute in relation to changes of the irrelevant attribute in high repetition conditions compared to low repetition conditions. One interpretation of these findings is that information about the irrelevant attribute (location or magnitude) in the previous trial is used much as an informative precue, so that participants can prepare early processing stages in the current trial, with the corresponding benefits and costs typical of standard cueing studies.

References

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do Fish count? Spontaneous discrimination of quantity in female mosquitofish. *Science*, **282**, 746–749.

Antell, S.E., & Keating, D.P. (1983). Perception of numerical invariance in neonates. *Child Development*, **54**, 695–701.

Arsalidou, M., & Taylor, M. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, **54**, 2382–2393.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two–choice responding task. *Quarterly Journal of Experimental Psychology*, **13**, 90–102.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467–472.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, **36**, 2257–2273.

Brannon, E.M., & Terrace, H.S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, **282**, 746–749.

Brannon, E.M., Wusthoff, C.J., Gallistel, C.R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a Linear Subjective Number Scale. *Psychological Science*, **12**, 238–243.

Buckley, P.B., & Gillman, C.B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, **103**, 1131–1136.

Dehaene, S. (2006). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, **8**, 47–68.

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

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

Ditz, H.M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of Royal Society B*, **283**, 20160083.

Festinger, L. (1943). Studies in decision. II. An empirical test of a quantitative theory of decision. *Journal of Experimental Psychology*, **32**, 411–423.

Fischer, M.H., Castel, A.D., Dodd, M.D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, **6**, 555–556.

Gold, J.I., & Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, **30**, 535–574.

Hassmann, M. (1952). Vom Erlernen unbenannter Anzahlen beim Eichhörnchen. *Zeitschrift für Tierpsychologie*, **9**, 294–321.

Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., & Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Frontiers in Human Neuroscience*, **5**, 186.

Henik, A., & Tzelgov, J. (1982). Is 3 greater than 5 – the relation between physical and semantic size in comparison tasks. *Memory & Cognition*, **10**, 389–395.

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

Holloway, I.D., Battista, C., Vogel, S.E., & Ansari, D. (2013). Semantic and perceptual processing of number symbols: Evidence from a cross-linguistic fMRI adaptation study. *Journal of Cognitive Neuroscience*, **25**, 388–400.

Huber, S., Moeller, K., & Nuerk, H.-C. (2014). Dissociating number line estimations from underlying numerical representations. *The Quarterly Journal of Experimental Psychology*, **67**, 991–1003.

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., Felber, S., & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *NeuroImage*, **25**, 888–898.

Livingstone, M.S., Srihasam, K., & Morocz, I.A. (2010). The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Animal Cognition*, **13**, 711–719.

Macmillan, N.A., & Creelman C.D. (2005). *Detection theory: A user's guide (2nd ed.)*. Mahwah, NJ: Erlbaum.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, **315**, 57–59.

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Animal Behaviour*, **47**, 379–387.

Moore, A.M., & Ashcraft, M.H. (2015). Children's mathematical performance: Five cognitive tasks across five grades. *Journal of Experimental Child Psychology*, **135**, 1–24.

Mou, Y., & vanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review*, **34**, 1–25.

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

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, **115**, 10–25.

Nieder, A., Freedman, D.J., & Miller, E.K. (2002). Representation of the quantity of visual Items in the primate prefrontal cortex. *Science*, **297**, 1708–1711.

Nuerk, H.–C., Moeller, K., Klein, E., Willmes, K., & Fischer, M.H. (2011). Extending the mental number line a review of multi-digit number processing. *Journal of Psychology*, **219**, 3–22.

Pepperberg, I. M. (1987). Evidence for conceptual quantitative abilities in the African parrot: labeling of cardinal sets. *Ethology*, **75**, 37–61.

Piazza, M., & Eger, E. (2016). Neural foundations and functional specificity of number representations. *Neuropsychologia*, **83**, 257–273.

Poltrock, S.E. (1989). A random walk model of digit comparison. *Journal of Mathematical Psychology*, **33**, 131–162.

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, **90**, 1392–1407.

Ratcliff, R., & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, **111**, 333–367.

Ratcliff, R., Smith, P.L., Brown, S.D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, **20**, 260–281.

Ratcliff, R., Thompson, C.A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy. *Cognition*, **137**, 115–136.

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

Rugani, R., Vallortigara, G., Priftis, K., & Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Animal Cognition*, **347**, 534–536.

Schwarz, W., & Eiselt, A.K. (2012). Numerical distance effects in visual search. *Attention, Perception & Psychophysics*, **74**, 1098–1103.

Schwarz, W., & H.–J. Heinze (1998). On the interaction of numerical and size information in digit comparison: A Behavioral and Event–Related Potential (ERP) Study. *Neuropsychologia*, **36**, 1167–1179.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number–size interference effects in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 507–522.

Schwarz, W., & Keus, I.M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, **66**, 651–664.

Schwarz, W. & Müller, D. (2006). Spatial associations in number-related tasks: A comparison of manual and pedal responses. *Experimental Psychology*, **53**, 4–15.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, **48**, 630–633.

Simon, J.R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, **81**, 174–176.

Smith, D.G., & Mewhort, D.J.K. (1998). The distribution of latencies constrains theories of decision time: A test of the random-walk model using numeric comparison. *Australian Journal of Psychology*, **50**, 149–156.

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**, 1324–1336.

Starkey, P., & Cooper, R.G. (1980). Perception of numbers by human infants. *Science*, **210**, 1033–1034.

Starkey, P., & McCandliss, B.D. (2014). The emergence of "groupitizing" in childrens numerical cognition. *Journal of Experimental Child Psychology*, **126**, 120–137.

Strauss, M.S., & Curtis, L.E. (1981). Infant perception of numerosity. *Child Development*, **52**, 1146–1152.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, **21**, 259–263.

Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, **10**, 130–137.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, **7**, 483–488.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, **358**, 749–750.

Xia, L., Emmerton, J., Siemann, M., & Delius, J.D. (2001). Pigeons (Columbia livia) learn to link numerosities with symbols. *Journal of Comparative Psychology*, **115**, 83–91.

Chapter 2

One model fits all: Explaining many aspects of number comparison within a single coherent model – A random walk account

Abstract

The time required to determine the larger of two digits decreases with their numerical distance, and, for a given distance, increases with their magnitude (Moyer & Landauer, 1967). One detailed quantitative framework to account for these effects is provided by random walk models. These chronometric models describe how number-related noisy partial evidence is accumulated over time; they assume that the drift rate of this stochastic process varies lawfully with the numerical magnitude of the digits presented. In a complete paired number comparison design we obtained saccadic choice responses of 43 participants, and analyzed mean saccadic latency, error rate, and the standard deviation of saccadic latency for each of the 72 digit pairs; we also obtained mean error latency for each numerical distance. Using only a small set of meaningfully interpretable parameters, we describe a variant of random walk models that accounts in considerable quantitative detail for many facets of our data, including previously untested aspects of latency standard deviation and error latencies. However, different from standard assumptions often made in random walk models, this account required that the distributions of step sizes of the induced random walks are asymmetric. We discuss how our findings can help in interpreting complex findings (e.g., conflicting speed vs. accuracy trends) in applied studies which use number comparison as a well-established diagnostic tool. Finally, we also describe a novel effect in number comparison, the decrease of saccadic response amplitude with numerical distance, and suggest an interpretation using the conceptual framework of random walk models.

Introduction

Comparing quantities is a fundamental cognitive skill of considerable survival value for animals and humans. Correspondingly, symbols to signal quantities in the abstract played a fundamental role in human cultural evolution, and humans are extremely efficient in comparing symbolic information representing distinct numerical magnitudes. Digit comparison involves "symbolic" information because the relevant dimension – the abstract numerical magnitude – cannot be directly inferred from the graphical form of the digits as such; rather, we have to rely on previously learned internal representations from which numerical information is retrieved.

In a classical study investigating the nature of these representations and retrieval processes Moyer and Landauer (1967) demonstrated that the time to choose the numerically larger out of two simultaneously presented digits systematically decreases with their numerical difference (the numerical distance effect; for further details, see, e.g., Dehaene, 1992, 2011; for general background, see Dehaene & Brannon, 2011; Nieder, 2005). The numerical distance effect is little influenced by extensive practice (e.g., Poltrock, 1989; Steinborn, Bratzke, Rolke et al., 2010), it has repeatedly been demonstrated with different number notation systems (e.g., Holloway, Batista, Vogel, & Ansari, 2013; Razpurker-Apfeld & Koriat, 2006; Takahashi & Green, 1983), it is observed in humans from early age on (e.g., Girelli, Lucangeli, & Butterworth, 2000; Holloway & Ansari, 2009; Landerl and Kölle, 2009; Moore & Ashcraft, 2015; Mussolin, Mejias & Noël, 2010; Sekuler & Mierkiewicz, 1978; Soltész, Szücs, Dékány et al., 2007), shows up consistently in different behavioral (e.g., Fischer & Miller, 2008; Fernandez, Rahona, Hervas, et al., 2011; Milosavljevic, Madsen, Koch, & Rangel, 2011; Song & Nakayama, 2008; Ganor-Stern & Goldman, 2015) and electrophysiological (e.g., Nieder, 2005; Schwarz & Heinze, 1998; Soltész et al., 2007; Cassey, Heathcote & Brown, 2014) indices, and has been demonstrated in numerous studies on animal cognition (e.g., Livingstone, Srihasam, & Morocz, 2010; Matsuzawa, 1985).

According to a prominent model originally proposed by Moyer and Landauer (1967; for summaries and elaborations, see Dehaene, 2011; Gallistel & Gelman, 1992, 2005; Nieder, 2005) symbolic information about numerical magnitudes is converted into percept–like analog representations which are then in turn compared to each other, much like the comparison of sensory representations of stimuli which differ along some extensive physical dimension. This analog representation is thought to be quickly accessible, fuzzy, and preverbal, and is often assumed to form a necessary developmental stage preceding the built–up of slower but exact verbal, propositional algorithms which form the basis of our abstract (e.g., algebraic) computational cognitive skills (e.g., Gallistel & Gelman, 1992, 2005; Hyde, Khanum, & Spelke, 2014; Norris, McGeown, Guerrini, & Castronovo, 2015; Starr, Libertus, & Brannon, 2013).

Convergent evidence supporting the concept of an analog representation of numerical magnitude comes from experimental, comparative, physiological, and developmental psychology (for detailed reviews, see Dehaene & Brannon, 2011; Nieder, 2005; Shettleworth, 2010, ch. 10). For example, one line of evidence consistent with the concept of an analog magnitude representation derives from Stroop-type interference experiments in which digits are presented in varying physical (font) sizes. Replicating and extending Besner & Coltheart (1979), many studies observed systematic number-size congruency effects, with both number and size as response-relevant dimensions (e.g., Henik & Tzelgov, 1982; Risko, Maloney & Fugelsang, 2013; Schwarz & Ischebeck, 2003), and these effects have well-documented neurophysiological correlates (e.g., Cohen-Kadosh, Cohen-Kadosh, Linden, Gevers, Berger, & Henik, 2007; Schwarz & Heinze, 1998). Similarly, the concept of a percept-like representation of numerical magnitude is often seen as consistent with related distance-dependent congruity effects observed in elementary perceptual tasks such as temporal order judgments (Schwarz & Eiselt, 2009), number naming in a flanker task (Ischebeck, 2003), or visual search (God-

win, Hout, & Menneer, 2014; Schwarz & Eiselt, 2012). Finally, like representations of sensory magnitudes, number comparison latencies exhibit a Weber-law signature: for a given numerical distance, the greater the numerical value of the smaller digit, the longer it takes to judge their order (the magnitude effect; Dehaene, 2003; Gallistel & Gelman, 2005; Nieder, 2005). In its strong version Weber's law predicts that latencies and error rates depend only on the ratio of the two digits involved. Note that the numerical distance effect is more general in the sense that Weber's law implies the distance effect but not conversely. Specifically, latency and error rate may well decrease monotonically with numerical distance, yet not in a manner conforming to Weber's law.

Random walk models of number comparison

Given this broad and multi–disciplinary interest in number comparison, surprisingly few attempts exist which seek to model quantitatively different aspects of representative number comparison data on the basis of a single coherent conceptual framework. Random walk and diffusion models (for general background, see Link, 1992; Luce, 1986, ch.s 8-9; Ratcliff & Smith, 2004) have considerable biological validity because they successfully account, at the single-cell level, for known neural correlates of two-choice decisions (e.g., Gold & Shadlen, 2007; Ratcliff, Cherian, & Segraves, 2003; Schall, 2001), and are often seen as a natural and parsimonious framework to incorporate the concept of analog magnitude representations.

The basic idea behind random walk models of number comparison (Figure 1) is that each of a pair of digits activates internal analog representations retrieved from memory. These representations provide noisy partial information about each digit which is contrasted in each processing step by means of a differencing rule, and this differential evidence is then accumulated over time. The running total (starting at $x = 0$) at any moment is an analog quantity representing and summarizing the total information retrieved from memory by the participant so far. An overt decision
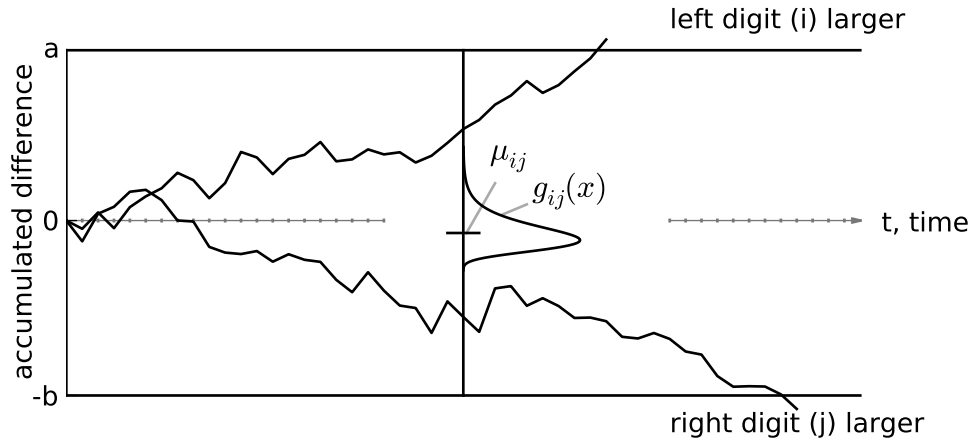
Figure 1. Random walk model of number comparison. Starting at $x = 0$, noisy partial evidence about the two digits (i, j) is accumulated until for the first time the upper response barrier at $x = a > 0$ or the lower response barrier at $x = -b < 0$ is reached. In the first case (upper sample path) the answer given is "left digit (i) larger", in the second case (lower sample path) "right digit (j) larger". A skew (double exponential) density $g_{ij}(x)$, illustrating the case $i < j$, is indicated from which the individual step sizes are drawn. The mean μ_{ij} of this density depends on the two digits presented; in the logarithmic model $\mu_{ij} = k \cdot [\ln(i) - \ln(j)]$, in the linear model $\mu_{ij} = k \cdot (i - j)$.

is reached when the cumulated differential evidence reaches for the first time an upper response barrier at $a > 0$ ("left digit larger") or a lower barrier at $-b < 0$ ("right digit larger"). Thus, the barriers represent the critical amount of differential evidence required by the participant for the particular response with which they are associated. For a given digit pair (i, j) response latencies depend on (the barriers and) the retrieval rates (the drifts μ_i, μ_j) associated with each of the digits; the net drift rate μ_{ij} of the accumulation process is then given by the difference of the individual drifts, $\mu_{ij} = \mu_i - \mu_j$. That is, the net drift rate μ_{ij} is the mean of the density $g_{ij}(x)$ associated with the steps of the random walk. For explicit analytic results for random walk models, see, e.g., Link (1975), Luce (1986, ch. 8), Schwarz (1990, 1991), or Townsend & Ashby (1983, ch. 10).

A number of studies has produced converging evidence suggesting that random walk or diffusion models (i.e., the continuous-time counterparts of discrete-time random walk models which result when increasingly smaller steps are made

at an increasingly higher speed; cf., Darling & Siegert, 1953; Feller, 1971; Luce, 1986; Smith, 1990) provide a coherent conceptual framework that is potentially able to account for basic results from a variety of important paradigms used in number comparison. Following early qualitative suggestions (Buckley & Gillman, 1974), Poltrock (1989) was the first to account for mean response time (RT) and error rates in a number comparison task quantitatively on the basis of a random walk model. Specifically, he showed that the effects of varying the emphasis on speed vs. accuracy are well accounted for by variations of the response barriers. In contrast, the digit-specific drift rates μ_i remained largely invariant across the speed conditions, and were an increasing, concave function of the numerical magnitude i. Smith and Mewhort (1998) confirmed and extended these findings by fitting a random walk model to RTs in a closely related paradigm in which participants judged single digits ($\neq 5$) to be smaller or larger than the standard of 5. Schwarz and Stein (1998) used a diffusion model to account in quantitative detail for the time-dependent pre-sampling effects of presenting one of the two digits with a short temporal head start. Schwarz and Ischebeck (2003) used the same model to explain why and to which degree numerical magnitude and (font) size effects interact when the digits are compared with respect to number vs. with respect to font size. Using the single-digit paradigm of Smith and Mewhort (1998), the diffusion model also successfully explained, on a distributional level, the interaction of numerical distance effects with target (a digit > 5) prevalence in a go-nogo task (Schwarz, 2001). Sigman and Dehaene (2005; also see Kamienkowski, Pashler, Dehaene, & Sigman, 2011) applied a single-barrier diffusion model in a study which employed a dual task paradigm (cf., Pashler & Johnston, 1998) involving a number comparison task as one component. On the basis of their diffusion model they accounted in considerable detail for task-2 processing delays arising at a central bottleneck stage, with number comparison both as the first or second task, varying numerical distance, notation, and response complexity as further factors.

A number of evidence accrual and accumulator models which are closely related to the basic random walk model have also been shown to account for a variety of systematic effects in number-related tasks (see, for example, Gallistel & Gelman, 2005; Whalen, Gallistel, & Gelman, 1999). Naturally, other important model architectures have also been proposed in the context of number comparison. For example, detailed simulation studies of complex computational models (e.g., Chen & Verguts, 2010; Moeller, Huber, Nuerk & Willmes, 2011; Verguts, Fias, & Stevens, 2005) have demonstrated unsupervised number-related learning effects in network units. These effects mimic several aspects in the processing of symbolic vs. non-symbolic stimuli, and can potentially help to differentiate spatial vs. non-spatial representations, or numerical vs. non-numerical order processing (e.g., Verguts & van Opstal, 2014) .

Open questions about random walk models of number comparison

Notwithstanding these encouraging results, the random walk accounts reviewed above have left open a number of important and potentially informative aspects of number comparison, reviewed in more detail below. This seems quite remarkable as random walk models provide a coherent chronometric framework, based on a dynamic, process-oriented point of view. Therefore, an important feature of these models is usually seen in their ability to address different aspects of number comparison data based on only a small set of elementary processing assumptions. We next discuss five of these aspects which we address in our study.

First, random walk models predict error rates but there is a clear lack of findings showing that these models account quantitatively for the full matrix of error rates obtained in a complete paired number comparison design. For example, in his extensive study Poltrock (1989, eq. 5) tested certain approximately linear overall relations between transformed error rates and mean overall RT for all pairs involving the digits from 2 to 9. From his regression analyses he concluded that the

random walk model accounted quite well for the major effects on error rate in his data. However, Luce (1986, p. 346) analyzed in detail overall linear regression techniques for random walk models (as also used by Poltrock, 1989) and warned that the conclusions suggested by these analyses "can be grossly misleading". To illustrate Luce's point, we substituted the parameter estimates given in Poltrock's Table 7 into his eq. (6) to obtain probability predictions for each individual digit pair separately. When these predictions are evaluated by a conventional χ^2-test for the individual binomial proportions (Poltrock, 1989, Table 6), the deviations are highly significant, $\chi^2(df = 41) \approx 9081$. Thus, although Poltrock's random walk model reproduces several features of the data such as the distance effect and the speed-accuracy trade–off quite well, its account of the individual error rates is not satisfactory (also see Poltrock, 1989, pp. 151-152). It is at present unclear if this lack of fit reflects a genuine, systematic inadequacy of the basic random walk model.

Second, the coherent chronometric framework provided by the random walk model allows in principle for prediction not only of mean RTs and error rates but also of RT variances, which provide a distinctive noise signature that reflects and ideally reveals basic aspects of the underlying processing mode (e.g., Whalen et al., 1999; Cordes, Gallistel, Gelman, & Latham, 2007). Variance predictions require no new assumptions to be added to the basic model: the variance of the decision times $\mathbf{D}_{ij}$ for a given digit pair (i, j) is completely defined by the processing assumptions summarized in Figure 1. Thus, using the standard additive decomposition $\mathbf{RT}_{ij} = \mathbf{D}_{ij} + \mathbf{M}$, the only new parameter needed (assuming independence of decision and motor times) is the variance of the non-decision component, $\mathbf{M}$. Despite this fact, no systematic attempt has so far been made to test whether random walk models can explain RT variance in number comparison tasks.

Third, a similar observation holds for error latencies as well. An attractive aspect of random walk models is their ability to predict mean RTs conditional on a correct

vs. incorrect response. For example, are errors in number comparison faster than correct responses? Or, is there a numerical distance effect for mean RT of incorrect responses as well, and if so does the random walk correctly account for it?

Most previous random walk accounts of number comparison either ignored the latency of errors altogether (e.g., Schwarz, 2001; Sigman & Dehaene, 2005, p. 347), or focused exclusively on overall RT, lumping together mean correct and error RT (e.g., Poltrock, 1989, eq. 5). The only related study which addressed the comparison of mean correct vs. error latency is that of Smith and Mewhort (1998) whose findings remained inconclusive. Their random walk model (p. 150) assumes normal (i.e., symmetric) step size densities and also symmetric boundaries (i.e., $a = b$); it predicts (cf., their eq.s 2–4) that error and correct RTs have the same distribution. In line with this prediction, these authors found (their Experiment 2; p. 155) that for a given digit mean error and correct RT did not differ. Because the pattern for error and correct RTs was similar, they did not model or analyze error latencies any further. In contrast, in their otherwise similar Experiment 1, errors tended to be faster than correct RTs (p. 152; their Table 2). Thus it is an open question whether error and correct latencies differ, and also whether error latencies in number comparison show a numerical distance effect.

More generally, random walk models for paired comparison designs typically assume or imply that the step size densities for complementary digit pairs (i, j) and (j, i) are mirror-images of each other: $g_{ij}(x) = g_{ji}(-x)$. As reviewed in the Appendix, if in addition the step size densities $g_{ij}(x)$ themselves are symmetric (such as the normal density; e.g., Smith & Mewhort, 1998, p. 150, or Poltrock, 1989, p. 138) then the mirror-image property implies (even when $a \neq b$) that incorrect responses to any given pair are on average as fast as correct responses to its complementary pair (note that these two cases lead to the same overt response). Averaging across both pairs (as is routinely done in computing the numerical distance effect), correct responses and errors are then predicted to be equally fast.

As an alternative to symmetric step size densities in the present study we consider the skew double exponential distribution (for background on this distribution in the context of perceptual decision making, see Foley & Schwarz, 1998; Schwarz & Miller, 2014) which is shown in Figure 1, and described formally in the Appendix. Critically, under the usual mirror-image property $g_{ij}(x) = g_{ji}(-x)$ this skew distribution does not predict errors for a given digit pair to be on average as fast as correct responses to its complementary pair. To see why, consider Figure 1 which refers to a case in which the left digit is smaller than the right digit (i.e., $i < j$). Negative steps are then directed towards the lower barrier at $-b$ which is associated with the response "right digit larger", and the random walk will on average tend to this correct barrier, i.e., $\mu_{ij} < 0$. Intuitively, a step size density $g_{ij}(x)$ with positive skew as shown in Figure 1 permits especially large isolated steps into the direction of its long (here: the positive) tail, leading to shorter mean first passage times across the (incorrect) barrier to which this long tail points. Thus, a skew step size density $g_{ij}(x)$ as in Figure 1 could explain from first principles why incorrect responses to a pair (i, j) are on average faster (Figure 1) or slower (i.e., when the elongated tail points towards the correct response barrier) than correct responses (j, i). For a more comprehensive and conclusive evaluation of random walk models of number comparison it would thus be important to know if the predictions implied by symmetric step size densities hold. If they do not hold then it would be equally important to know if this reflects a genuine failure of the random walk model in Figure 1 per se (necessitating additional modifications, such as, e.g., starting point variability across trials; see Brown & Heathcote, 2008; Donkin, Brown, Heathcote, & Wagenmakers, 2011; Ester, Ho, Brown, & Serences, 2014; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004) or else if skew (e.g., double exponential) distributions are able to provide a more adequate and complete account of the data.

Fourth, there has been an extensive debate (e.g., Cordes et al., 2007; Dehaene, 2003, 2011; Huber, Moeller & Nuerk, 2014; Nieder, 2005) about the na-

ture and scale (e.g., linear vs. logarithmic) of the mapping of numerical magnitude onto its internal representation, and how the functional form of this mapping relates to the Weber-law signature of numerical comparison latencies. Random walk models as in Figure 1 provide a well-defined framework which gives a clear operational meaning to these conceptual distinctions. Specifically, in the context of random walk models a crucial aspect is the manner in which the mean size of each step in the accumulation process (i.e., the net drift rate) depends on the digits presented. Most extant applications of random walk models of number comparison have treated and estimated these drift rates as free, unconstrained parameters, and have not specifically addressed or compared different functional forms of number-to-drift mappings. In the present study we aimed to compare linear vs. logarithmic number-to-drift mappings within the conceptual framework of random walk models, which also provides a theoretical basis to explore the open question of analogous Weber-law signatures for error rates and the variability of latency.

Finally, as random walk models focus on decision processes, we intended to minimize the contribution of non-decisional processing stages, $\mathbf{M}$. For example, using manual responses Poltrock (1989, Tables 1 and 4) obtained a mean non-decisional component of $\mu_M = 378$ ms for mean RTs of about 420 ms. According to this estimate, with button press responses non-decisional processing stages make up about 90 per cent of total RT (for similar results using manual responses, see Sigman & Dehaene, 2005; Smith & Mewhort, 1998). In contrast, our participants indicated their decisions by horizontal saccades which are well-known to be faster and less variable than manual responses (e.g., Khalid & Ansorge, 2013, Figures 5 and 6); for example, in their number comparison study Milosavljevic et al. (2011) estimate that when their participants just exceeded the level of chance performance then their saccades had onset latencies of only about 230 ms. Thus, while saccadic responses do not eliminate the need to program and execute an overt motor response, they serve to reduce the contribution of the motor compo-

nent (e.g., its mean and variance) relative to the decision component. In addition, measuring saccadic eye movements enabled us to investigate if other aspects of the response beyond its accuracy and latency (e.g., saccadic amplitude) are systematically related to numerical distance as well (e.g., Fischer & Miller, 2008; Song & Nakayama, 2008; Ganor-Stern & Goldman, 2015).

Conclusions regarding the issues reviewed above can have potentially relevant implications for applied studies which use number comparison as a well-established diagnostic tool in currently active research areas (e.g., Norris et al., 2015). For example, as described in more detail in our General Discussion, in some of these studies, speed and accuracy measures work in opposite directions when different groups of participants (e.g., young vs. elderly adults) are compared. If random walk models do provide an adequate framework to account for various aspects of number comparison then they would seem ideally suited to disentangle opposing speed and accuracy trends, and therefore could help to interpret complex findings.

Thus, the aim of the present study was to address the five topics reviewed above in a complete paired number comparison design with saccadic responses, using a random walk framework (Figure 1) with a minimum of parameters. More specifically, within this framework we wanted to predict mean correct saccadic onset latency (SL) as well as error rates and SL standard deviations on the level of the 72 individual digit pairs. Also, we aimed at testing if (as predicted by random walk models) error responses, too, show a numerical distance effect, and to compare systematically mean correct to mean error SL of complementary digit pairs. These comparisons test the strong predictions made by random walk models with symmetric (e.g., normal) step size densities, and evaluate the plausibility of accounts based on skew step size densities. Finally, using the framework of random walk models we aimed at comparing different mappings (linear vs. logarithmic) of numerical magnitudes onto their internal representations which govern the drift rates of the evidence accumulation process.

Method

Participants. Forty-three University Potsdam students, aged 18-33, with normal or corrected-to-normal vision participated in one session of approximately one hour. They either received a payment of €8 or course credit for their participation.

Stimuli and apparatus. The stimuli consisted of all 72 ordered digit pairs (i, j), $1 \leq i, j \leq 9$. Each display consisted of two different digits side by side to the left and to the right of a centered fixation cross on a 75-Hz, 1028 $\times$ 768 pixel (px) VGA color monitor; the display timing was synchronized with the video refresh cycle. The horizontal distance between the centers of the digits was 30 px (from the viewing distance of the participants of 114 cm, 1 deg = 65 px). The digits were presented in dark blue against a gray background in Verdana font using 24 px font height. The participants responded with saccades to two permanently visible square boxes (50 $\times$ 50 px), with their centers 256 px to the left and 256 px to the right of the fixation cross. In the center of both square boxes was a red dot with a diameter of 6 px.

The SMI Iview-X was used to track the left eye, using a sampling rate of 240 Hz. To detect the on- and offsets of saccades, a saccadic velocity peak threshold of 75 deg/sec was used. A chin- and headrest was used throughout the experiment.

Procedure. In each of four blocks all 72 ordered digit pairs were presented once in random order. The task was to indicate, as quickly as possible, the larger digit. The required response was a saccade to the center point of the square box into the direction of the larger digit. Each trial started with the presentation of a fixation cross. After a random delay of 700-800 ms the two digits were presented. A blank screen was presented after 2000 ms for 500 ms, before the next trial started.

Preliminary Data Reduction. For a trial to be valid the central cross had to be fixated at least from the presentation of the digits until the onset of the response saccade. In addition, the horizontal distance between the fixation cross and the

response saccade's landing point was required to be within 128 px and 512 px, and its latency (time from the onset of the digits to the onset of the saccade) between 120 ms and 900 ms. Invalid trials not meeting these criteria were excluded from all further analyses. There were six per cent invalid trials in the whole data set of $43 \times 288 = 12384$ trials (four per cent due to technical tracking problems, one per cent due to a failure to fixate the central cross, and one per cent due to SL outliers).

For each participant and each digit pair mean correct SL, error rate, standard deviation (SD) of correct SLs, and mean amplitude of the response saccades were computed from all valid trials. Saccadic response amplitude was defined as the horizontal distance between the fixation cross and the landing point of a response saccade.

Results

Numerical Distance Effect. Numerical distance had a main effect on SL ($F(7, 294)$ = 74.26, MSE = 248.59, p < .001, η^2 = .64); the linear trend showed that mean SL decreased by 8.5 ms per distance unit on average across participants ($t(42)$ = -13.48, p < .001, η^2 = .68). Similarly, the SD of SL decreased with numerical distance ($F(7, 294)$ = 26.91, MSE = 165.48, p < .001, η^2 = .39), on average by 4.4 ms per distance unit ($t(42)$ = -4.86, p < .001, η^2 = .22). Mean error rates for each numerical distance were subjected to a logistic regression. At numerical distance 1 the error odds were .04 (corresponding to an error rate of 3.85 per cent), they decreased by a factor of .63 per distance unit ($t(6)$ = -3.89, p < .01, η^2 = .52).

Random Walk Model: Test of symmetry prediction. As reviewed in the Introduction, a basic prediction of random walk models with symmetric step size densities is that mean SL for error responses to a given digit pair (i, j) should be equal to the mean SL for correct responses to its complementary digit pair, (j, i). To test this fundamental property, we compared each individual error SL with the correspond-
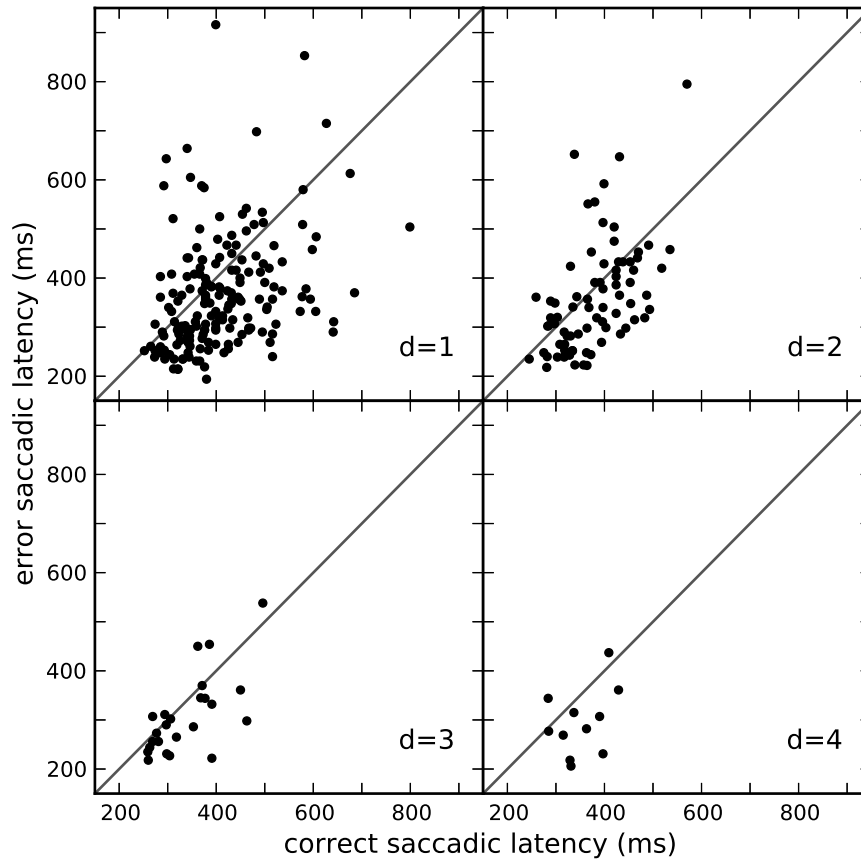
Figure 2. We computed for each individual error saccadic latency (SL) for a given digit pair (i, j) the mean correct SL of the complementary digit pair (j, i) of the same participant. Error SLs (ordinate) were plotted against their associated complementary mean correct SLs (abscissa) for different numerical distances (d). Upper left panel (d=1): 127 of 182 (69.7 per cent, p < .001, η^2 = 0.88), upper right panel (d=2): 53 of 79 (67.1 per cent, p < .01, η^2 = 0.76), lower left panel (d=3): 19 of 24 (79.2 per cent, p < .01, η^2 = 0.71), and lower right panel (d=4): 9 of 11 (81.8 per cent, p = .065, η^2 = 0.58) data points fall below the main diagonals, indicating systematically faster error responses at each numerical distance; all comparisons by sign tests.

ing mean correct SL of its complementary pair; we emphasize that all 307 individual comparisons were carried out within a given participant, without pooling data across subjects. For example, if a participant made an error in response to digit pair $(7, 8)$, we compared this error SL to the mean correct SL of this participant for the pair $(8, 7)$. Overall, 214 of 307 (69.7 per cent) incorrect responses were faster than their complementary mean correct responses, a rate that differs substantially from 50 per cent (p < .001, sign test, η^2 = .93). In some paradigms, error re-

sponses switch from being faster than correct responses to being slower than correct responses, when the decisions are made more difficult (see, e.g., Luce, 1986; Ratcliff & Rouder, 1998). Therefore, we tested the symmetry prediction separately at different levels of task difficulty (i.e., at different numerical distances). As shown in Figure 2, these tests revealed that incorrect responses were faster than their complementary mean correct responses for each individual distance tested, and not just for particularly easy comparisons. To further investigate the consistency of the effect shown in Figure 2, we also compared for each participant separately how many of his/her individual error SLs were faster/slower than the corresponding mean correct SL of the complementary digit pair. Only 6 of 41 participants (2 participants committed no error at all) produced more incorrect responses that were slower (rather than faster) than the mean SL of correct responses to the complementary digit pair (p < .001, sign test, η^2 = .87). These results confirm that the main finding of faster errors to (i, j) compared to correct responses to (j, i) shown in Figure 2 holds consistently across participants, and cannot simply be attributed to a few observers contributing disproportionately many errors. Taken together, these analyses provide strong evidence that at all numerical distances incorrect responses to a pair (i, j) are consistently faster than correct responses to the complementary pair, (j, i). This robust finding is inconsistent with the predictions of any random walk model that assumes a symmetric step size density. Therefore, we next tested if a random walk model with a skewed (double exponential) step size density can account more adequately for our findings.

Random Walk with asymmetric step size densities: The double exponential model

To fit the random walk model with double exponential step sizes, we first evaluated its ability to account for the averaged data across participants; fits of the model to the data of each participant individually will be compared to the fit of the averaged data when we evaluate the model's account of error latencies. Thus, we defined an objective function f that compared for each digit pair observed mean

correct SLs and mean error rates of all 72 digit pairs with the corresponding predic-
tion of the model; a formal description of f is given in the Appendix. Specifically, for
each digit pair, correct SL and error rate were first averaged over all participants.
Next, we divided the difference between predicted and observed mean SL by its
standard error (s.e.), and squared this quantity to obtain an approximate $\chi^2(1)$.
Similarly, the difference between each observed and predicted error rate was eval-
uated by a $\chi^2(1)$ component in the standard manner. The two approximate $\chi^2(1)$
values were then summed, and aggregated across all 72 digit pairs; the objec-
tive function f was then numerically minimized in its four arguments: a, b, k, μ_M.
As reviewed in the Introduction (Figure 1) and in the Appendix, here a and b are
the response barriers of the random walk process, and the scale parameter k de-
scribes how the individual drift rates μ_i, μ_j of the two digits map onto the net drift
rate of the random walk process. In the version of the model that we first fitted to
our data set, we assumed a logarithmic mapping, $\mu_{ij} = k \cdot [\ln(i) - \ln(j)]$. Finally,
μ_M represents the mean non-decisional SL component.

The best fit of the double exponential model is shown in Figure 3. The response
barriers were $a = 13.69$ for responses to the left and $b = 14.06$ for responses
to the right (cf., Figure 1). From the best-fitting scale parameter $k = 0.907$ we
computed the net drift rate μ_{ij} for each digit pair (i, j). For example, for the digit pair
$(i = 6, j = 8)$ the net drift rate $\mu_{ij} = k \cdot [\ln(i) - \ln(j)] = 0.907 \cdot [\ln(6) - \ln(8)] = -0.26$;
the density with this drift rate is shown in Figure 1 and in Figure 7 in the Appendix.
The non-decision component μ_M means that on average 347 ms of the SL was
not involved in the decision process. Thus, with only four parameters we predicted
SLs and error rates for all 72 digit pairs, a total of 144 data points.

Figure 3 illustrates that the model accounted quite well for SLs (left panel) and
error rates (middle panel). As indicated by error bars in the left panel of Figure
3, most SL predictions fall within the 95 per cent confidence interval (CI) of mean
observed SL. The root mean squared deviation (11.6 ms) between predicted and
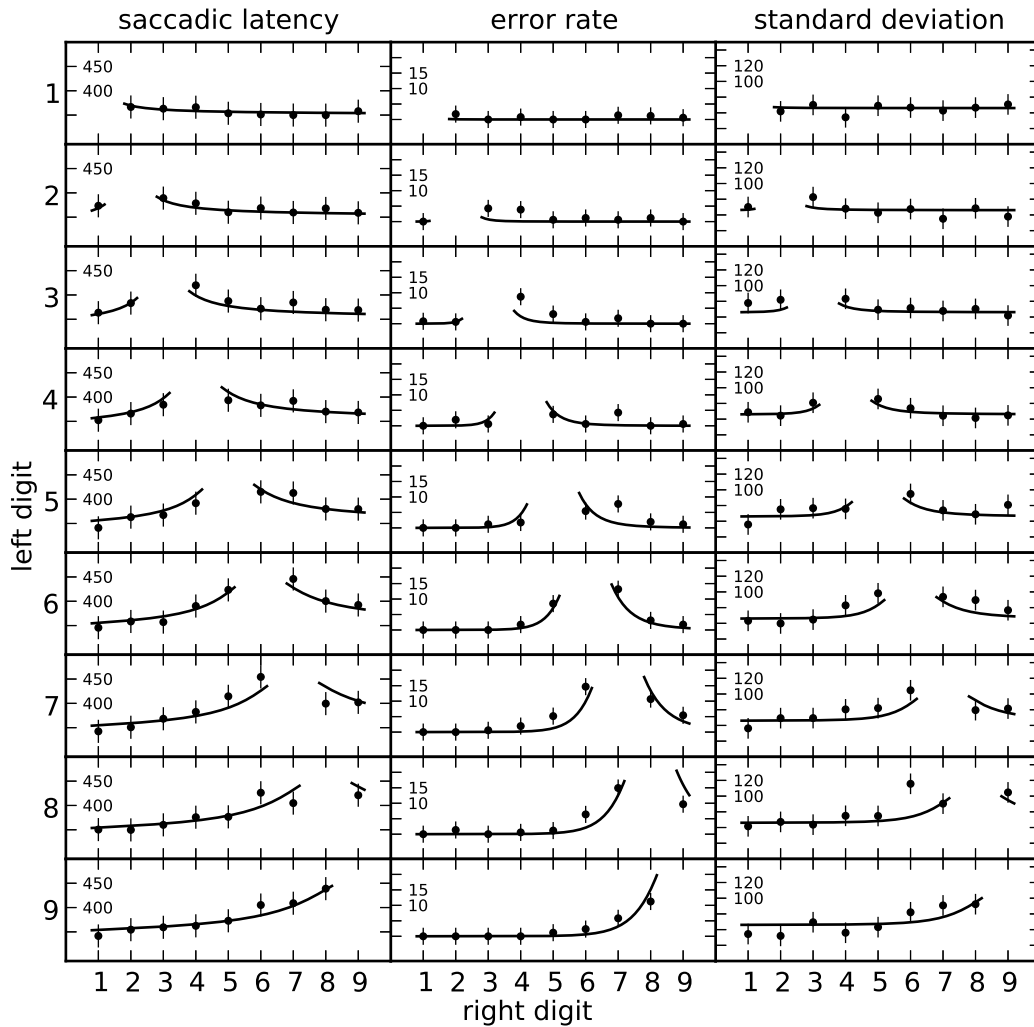
Figure 3. Mean correct saccadic latencies (SL), mean error rates, and mean standard deviation for all 72 digits pairs. Each sub-panel shows data and prediction for a given left digit (leftmost ordinate) and all eight right digits (abscissa). The inner ordinate for each sub-panel is: SL (ms) for left panel, error rate (per cent) for middle panel, standard deviation (ms) for right panel. Points show the data and solid lines show the model predictions. Error bars represent ±2 standard errors (Loftus & Masson, 1994).

observed SLs for the 72 digit pairs corresponded closely to the estimated standard error (11.2 ms) of the observed SLs across the participants. A similar finding held for the error rates: as indicated in the middle panel of Figure 3, most predicted error rates fall within the 95 per cent CI of mean observed error rate.

One important aspect of any model of number comparison is its ability to pro-
vide a coherent account of different aspects of the comparison process. To further
validate the model, we tested it with a dependent variable that was not itself in-
volved in the parameter estimation: the standard deviation (SD) of the SL. The
way in which the variability of SLs varies across the digit pairs provides a distinc-
tive noise signature reflecting basic aspects of the underlying process. Variance
predictions require no new assumptions to be added to the model (Schwarz, 1991):
the variance of the decision times (D) is completely defined by the basic random
walk process. In the random walk model SL consists of a decision component and
a non-decision component (SL $=$ **D** $+$ **M**). As a first approximation, we assumed
that **D** and **M** were independent, so that the variance of SL was the sum of the
variances of the decision and the non-decision component ($\sigma_{SL}^2 = \sigma_D^2 + \sigma_M^2$). Thus,
to predict the SDs only one additional model parameter (σ_M^2) was needed.

For any given digit pair, the observed variances were first averaged across all
participants. Next, using a least-square criterion, we estimated the variance of the
non-decision component, yielding $\sigma_M^2 = 66^2$. To obtain the predicted SD for each
digit pair we computed σ_D^2 from the model (Schwarz, 1991) with the best-fitting
parameters ($a = 13.69$, $b = 14.06$, $k = 0.907$), which had been estimated using only
the mean SLs and the error rates. As indicated by error bars in the right panel
of Figure 3, the SD predictions are quite close to the observed data, with most
predictions falling within the 95 per cent CI of mean observed SD. Regarding these
predictions, we stress that the empirical SDs of SL had not been used in any way
to find the original model parameters from which the predictions in the right panel
of Figure 3 were then generated.

Number-to-drift mapping.   As noted in the Introduction, another attractive fea-
ture of random walk models is that they permit testing different assumptions about
how the internal representation of the individual digits map onto the net drift rate
of the induced random walk process. In the model fit described above we used

a logarithmic mapping of individual digit representation onto the net drift rate, $\mu_{ij} = k \cdot [\ln(i) - \ln(j)]$. The best fit resulted in $f(a, b, k, \mu_M) = 181$ for 144 independent data points, which formally corresponds to a p-value of 2.0 per cent, and indicates that the logarithmic model is roughly in line with the $\chi^2(df = 144)$ distribution with an expectation of 144 and an SD of 17. To explore alternative mappings of individual drifts onto net drift rates, we also minimized the objective function using a linear mapping, $\mu_{ij} = k \cdot (i - j)$, under otherwise identical assumptions. However, when compared to the logarithmic model shown in Figure 3, the linear number-to-drift mapping resulted in a much poorer model fit, $f(a, b, k, \mu_M) = 416$, which formally corresponds to $p < .001$ and indicates much larger discrepancies between data and model predictions than would be expected if the linear model were correct. Overall, this comparison suggests that the logarithmic model represents a more adequate description of how individual digit drifts map onto the net drift rate.

How reliable and powerful are these random walk model comparisons between the logarithmic and the linear number-to-drift mapping? The large difference obtained for the best fits to our data under both models suggests that our design and procedure had some power to identify the more adequate model, but it remains unclear to which degree this may be attributable to random variation across hypothetical replications. To investigate this aspect more systematically we simulated all 288 individual trials of all 43 participants ten thousand times under the logarithmic and ten thousand times under the linear random walk model with double exponential steps on the basis of the parameter values (including $\sigma_M = 66$) reported above. Between-participant variability was added by drawing the model parameters for each participant separately from a normal distribution with mean equal to the values reported above, and a standard deviation of 20 per cent of that mean. For each individual simulation we then fitted both models to the simulated data, using the objective function described above and more formally in the Appendix.

When the data were generated using the logarithmic model the linear model fitted better in only 9 (of 10000) cases, and with data generated from the linear model the logarithmic model fitted better in only 14 (of 10000) cases. These results clearly suggest that our experimental design and model fitting procedure had considerable power to discriminate at least between relatively broad model alternatives such as random walk models based on a logarithmic vs. linear number-to-drift mapping.
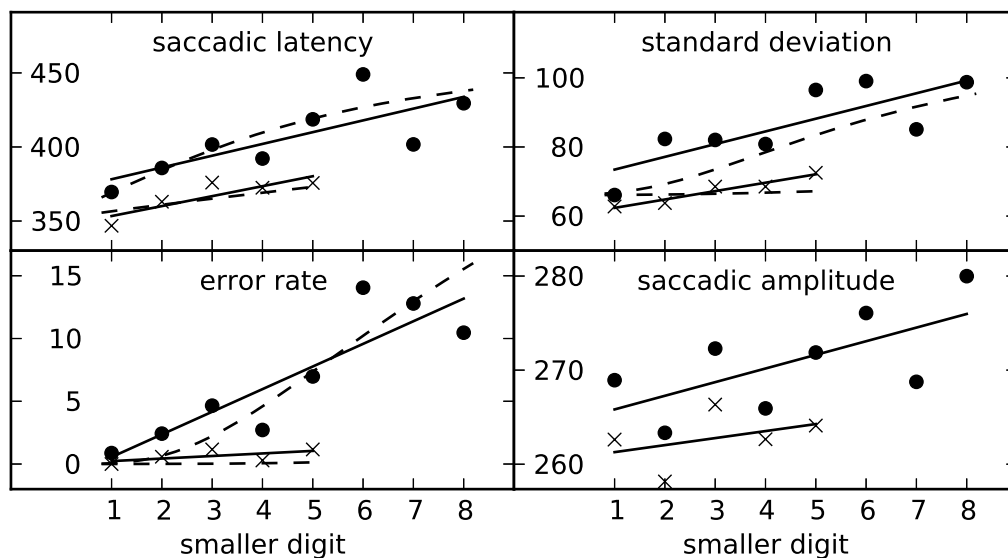

Figure 4. Weber-law signatures in number comparison. In all four panels, the smaller number of a digit pair is plotted on the abscissa; the curve parameter refers to a numerical distance of 1 (circles) or 4 (crosses). Top left panel: symbols show mean saccadic latency (SL) (ordinate in ms); the solid lines are the linear regression of mean SL on the smaller digit for each distance separately, the dashed lines show the corresponding predictions of the double exponential random walk model. Top right panel: same as top left panel, for the standard deviation of SL (ordinate in ms). Lower left panel: same as top left panel, for error rate (ordinate in per cent). Lower right panel: symbols show saccadic amplitude (ordinate in px), the solid lines are the linear regression of saccadic amplitude on the smaller digit for each distance separately.

The conclusions reached above are further supported by an analysis of the magnitude, or Weber-law, effect which is closely related to the logarithmic model. We first investigated for all three dependent variables (mean and SD of SL, and error rate) separately the presence and nature of the empirical magnitude effect.

Figure 4 illustrates for small ($d = 1$) and large ($d = 4$) numerical distance pairs that all three variables exhibit highly systematic Weber-law signatures in which for a given numerical distance mean SL, error rate, and the SD of SL all increase with numerical magnitude. We next asked whether the double exponential random walk model with logarithmic mapping quantitatively accounts for all effects simultaneously, using the single set of parameter estimates described above. Figure 4 indicates that the magnitude effect for all three dependent variables individually is well captured by the double exponential random walk model.

Table 1. Digit pairs with common ratios.

pairs involved	mean SL (s.e.)	SD of SL (s.e.)
(1,4), (2,8)	359 (9), 359 (9) F(1,42)=0.01 p=.93, η^2<0.01	56 (5), 59 (5) F(1,42)=0.40 p=.53, η^2=0.01
(1,3), (2,6), (3,9)	363 (11), 363 (11), 364 (10) F(2,84)=0.01 p=.98, η^2<0.01	64 (6), 57 (4), 58 (5) F(2,84)=0.95 p=.39, η^2=0.02
(1,2), (2,4), (3,6), (4,8)	370(12), 372(10), 364(10), 373(11) F(3,126)=0.73 p=.54, η^2=0.02	61(4), 59(4), 61(5), 60(5) F(3,126)=0.02 p=.99, η^2<0.01
(2,3), (4,6), (6,9)	368 (11), 368 (11), 398 (9) F(2,84)=3.25 p=.04, η^2<0.07	72 (6), 71 (5), 71 (5) F(2,84)=0.02 p=.98, η^2<0.01
(3,4), (6,8)	402 (12), 413 (11) F(1,42)=2.56 p=.12, η^2=0.06	73 (6), 93 (7) F(1,42)=8.80 p=.01, η^2=0.17

Note. Comparison of means and standard deviations (SD) of saccadic latency (SL) for digit pairs having common ratios. Numbers in brackets indicate the standard error (s.e.) of the mean or SD. Results averaged across complementary pairs such as $(1, 2)$ and $(2, 1)$.

A strong related prediction of the logarithmic mapping model is that equal digit ratios should produce equal performance, whereas the linear mapping model predicts that equal digit differences produce equal performance. For example, the net drift rate $\mu_{ij} = k \cdot \ln(i/j)$ (and thus performance) should be the same for the equal-ratio pairs (2,1), (4,2), (6,3), and (8,4), and similar predictions can be tested involving several other digit pairs. Table 1 shows results for comparisons involving all equal-ratio pairs, aggregated across complementary pairs, to increase statistical power. These results show that, as predicted by the logarithmic mapping model, pairs characterized by equal ratios produced means and (with one exception) SDs of SL which do not differ significantly for a given ratio.[1] We emphasize that the p values in Table 1 cannot be used to measure the strength of the support for the Weber model but simply indicate if the data differ from the Weber-law predictions by more than can be accounted for on the basis of random error alone.

Mean SL of incorrect responses. Do error responses show a numerical distance effect? As reviewed in the Introduction, evidence on this point is scant and ambiguous. According to the double exponential random walk model errors reflect a noisy process of weighing up partial evidence that is characterized by a small signal-to-noise ratio, and it predicts that there should be a numerical distance effect for incorrect responses as well. As a further validity test we compared this prediction to our data and analyzed mean error SL as a function of numerical distance, $|i - j|$.

Overall, errors were infrequent: a rate of 2.6 per cent (307 trials) across all digit pairs and participants. Numerical distances larger than 4 contributed altogether only 11 responses (3.6 per cent of all error responses). Therefore, only numerical distances from 1 to 4 were used for further analysis; the means from these 296 error SLs are shown in Figure 5. Mean error SL decreased on average by 25.0 ms

---

[1]Too few errors were observed for the pairs involved to test the corresponding model prediction for error rate as well.

per distance unit (t(294) = -2.87, p $<$ .01, η^2 = .11), indicating a robust numerical distance effect for error SLs.

To further validate this finding, we also considered error SL as a function of numerical distance for each participant individually. We computed the individual regression slope of mean error SL against numerical distance for each participant (a total of 33) who committed errors at least at two different distances. A negative regression slope was obtained for 28 of 33 participants (p $<$ .001, sign test, η^2 = .84), again suggesting that mean SL decreases with numerical distance, even if the response given was incorrect.

These results represent clear evidence for a numerical distance effect on error SLs, as predicted by random walk models. We next considered to which degree the double exponential random walk model accounts quantitatively for individual error SLs. More specifically, we analyzed error SL as an additional dependent variable, and tried to predict it with that model, using parameters that were first estimated exclusively from mean correct SL and error rate, i.e., without using in any way empirical error SLs.

To this end, we first fitted, in the same way as was done with the means across participants, the double exponential random walk model for each participant individually to estimate his/her individual model parameters. For most participants, the fit of the objective function f was quite good, as judged by small or moderate values of the statistic f_{min}. The medians of the parameters from these 43 individual model fits are: $a = 13.15$, $b = 14.90$, $k = 1.055$ and $\mu_M = 327$. These median parameters are all remarkably close to the parameters as obtained from the model fit to the means across all participants. Next, predicted error SLs of a given participant were computed for each digit pair for which that participant had made at least one error (altogether 296 predicted error SLs). Predicted error SLs were then subjected to a regression analysis, in the same way as was reported above for the observed error SLs. Predicted error SL decreased on average by 28.0 ms per
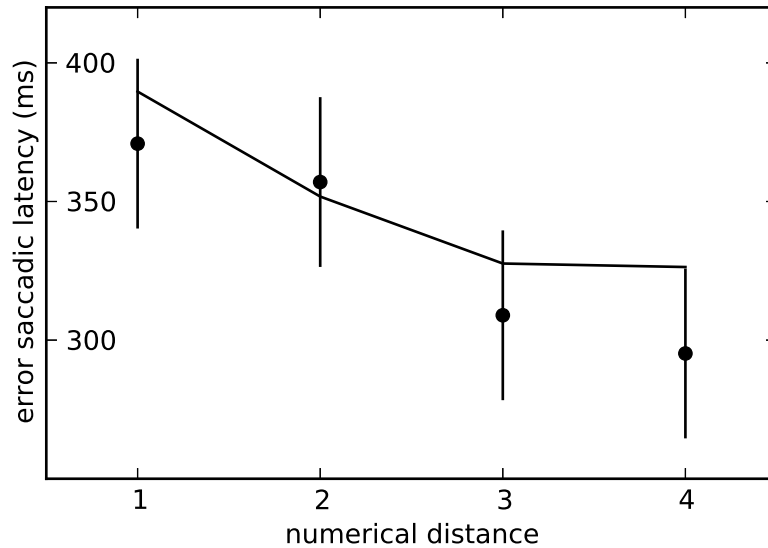
Figure 5. Error saccadic latencies (SL) as a function of numerical distance. Ordinate: error SL (ms). Abscissa: numerical distance, $|i - j|$. Data points show mean observed error SL, error bars show ± 2 standard errors, and solid line shows mean predicted error SL.

distance unit (t(294) = -5.55, p $<$ .001), a slope that corresponds closely to that (25.0 ms) obtained from the corresponding regression analysis of the observed error SLs. Finally, we also compared predicted and observed error SLs directly. In this analysis, the root mean squared deviation between predicted and observed error SLs equaled 6.4 ms; deviations of this magnitude fall well within ± 2 s.e. of the observed error SLs averaged for each numerical distance, as illustrated in Figure 5.

Saccadic Overshoot: A new measure of the numerical distance effect. Does numerical distance also affect other aspects of the response, beyond its correctness and latency? Several findings in the literature suggest that numerical distance may influence specific aspects of motor behavior. For example, the force of manual responses increases with increasing numerical distance (Fischer & Miller, 2008, Exp. 1), and response trajectories of manual pointing responses are more bent as numerical distance decreases (Song & Nakayama, 2008; Ganor-Stern & Gold-

man, 2015). To investigate this aspect for eye movements we analyzed saccadic response amplitude as a function of numerical distance. As shown in Figure 6, saccadic amplitude decreased systematically with increasing numerical distance ($F(7, 294)$ = 8.37, MSE = 127.85, p $<$ .001, η^2 = .17); on average by 1.5 px per distance unit (t(42) = -4.57, p $<$ .001, η^2 = .20). This means that on average, as numerical distance decreased participants responded with "overshooting" saccades showing more eccentric landing positions. To further validate this finding, we also analyzed the magnitude effect for saccadic amplitude. Figure 4 (lower right panel) illustrates for pairs with numerical distances of $d = 1$ and $d = 4$ that, for a given numerical distance, saccadic amplitude systematically increases with numerical magnitude.
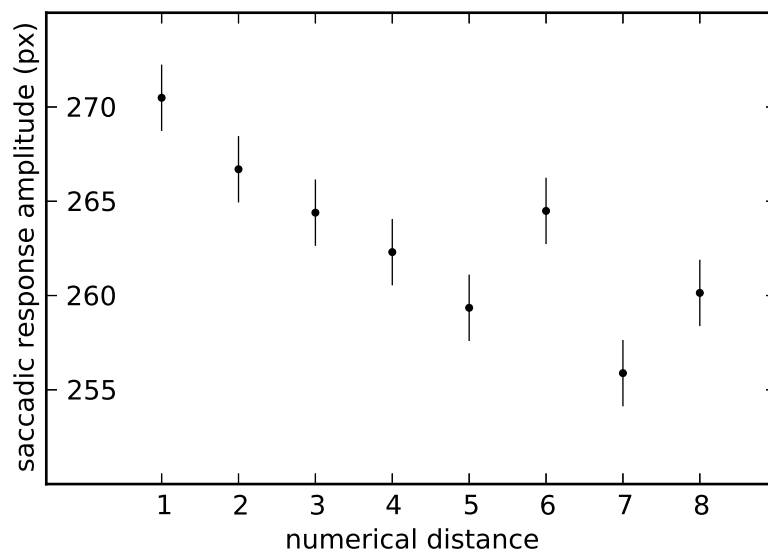
Figure 6. Mean saccadic response amplitude as a function of numerical distance. Ordinate: saccadic response amplitude (px). Abscissa: numerical distance. Data points show mean saccadic response amplitude, error bars show ± 1 standard error (Loftus & Masson, 1994).

Given that saccadic response amplitude was defined in absolute coordinates, i.e., by the distance of the landing point relative to the fixation cross (screen center), this finding might possibly reflect systematic distance-related saccade start-

ing point differences. Specifically, participants might start their saccades slightly more eccentrically with smaller distances, but then execute saccades of similar amplitude. Thus, we analyzed an alternative relative measure of the response amplitude, namely the horizontal distance between the actual starting point (rather than the fixation cross) and the landing point of a saccade. With this alternative measure, we confirmed the basic amplitude findings reported above. Specifically, horizontal distance decreased with numerical distance ($F(7, 294)$ = 9.82, MSE = 133.05, p < .001, η^2 = .19); on average by 1.6 px per distance unit ($t(42)$ = -5.13, p < .001, η^2 = .24). We conclude that the new finding of the saccadic overshoot effect does not simply reflect distance-related saccade start point differences, and occurs similarly for both measures of response amplitude.

Is this saccadic overshoot effect mediated by SL? As reported above, participants need more time to respond to digit pairs with small numerical distances, and this might in turn allow for a more effective preparation, leading to more eccentric saccades. For example, in a double step task Becker and Jürgens (1979) observed increasing saccadic amplitudes with increasing preparation time. Similarly, Thura, Cos, Trung, and Cisek (2014; Figure 5c) report that the saccadic amplitudes of macaque monkeys performing a visual choice task systematically increased with the time they needed to reach a decision, as predicted by their urgency gating account. On the other hand, if the overshoot effect is not mediated by SL, then saccadic amplitude should not be influenced systematically by differences in SL. Therefore, we compared the saccadic amplitude effect for faster vs. slower SLs. Specifically, for each participant and each digit pair separately, saccadic amplitudes were median-splitted into fast vs. slow responses.

In a repeated-measures ANOVA with the factors numerical distance (1 - 8) and SL (slow vs. fast) only numerical distance had an effect on saccadic amplitude ($F(7, 294)$ = 4.22, MSE = 407.85, p < .001, η^2 = .09). Neither slow vs. fast SL ($F(1, 42)$ = .79, MSE = 876.54, p = .38, η^2 = .02) nor the interaction ($F(7, 294)$ =

1.34, MSE = 297.52, p = .23, η^2 = .03) showed any effect on saccadic amplitude. Thus, the saccadic overshoot effect is observed similarly for slow vs. fast responses to a given digit pair. Possible explanations of this effect in terms of the random walk model will be examined in the General Discussion.

General Discussion

As reviewed in the Introduction, random walk and diffusion models have in previous studies quite successfully, and in considerable quantitative detail, accounted for results from many different experimental paradigms used in numerical cognition, including number-size congruency effects (Schwarz & Ischebeck, 2003), dual-tasks involving number comparison (Sigman & Dehaene, 2005; Kamienkowski et al., 2011), SOA-dependent effects of delaying one digit in a comparison pair (Schwarz & Stein, 1998), the speed-accuracy trade-off in numerical comparison (Poltrock, 1989), or single-digit comparison to a fixed numerical standard (Smith & Mewhort, 1998; Schwarz, 2001). A starting point of the present work was to ask if, under which assumptions, and to which degree, random walk models can also account for various more detailed facets of paired number comparison performance in humans. Specifically, in addition to standard measures such as the mean latency of correct responses, we focused on such aspects as error rates, latency variance, and the effect of numerical distance on the latency of error responses. To minimize non-decisional contributions to overall latency we used eye movements as responses (e.g., Khalid & Ansorge, 2013; Milosavljevic et al., 2011), and aimed at an account at the level of all 72 individual pairs made up of the digits 1–9.

When humans compare numbers, a standard assumption is that they convert the numerical magnitudes into an internal mental analog representation, followed by a comparison process much like that studied in psychophysics using extensive physical attributes, such as luminance or sound pressure. As Chen and Verguts

(2010) observed, the exact nature of this internal representation, sometimes supposed to be quasi-spatial in nature, "has remained elusive, however". Random walk models as in Figure 1 provide one means of making more specific, in a quantitative form, the notion and meaning of these representations. Informed by studies of basic neuronal mechanisms of magnitude coding (e.g., Nieder, 2005) these models assume that humans retrieve noisy representations of the difference of the numbers presented, and accumulate this fuzzy partial evidence until one of two evidence boundaries is reached for the first time. Although in principle all of the various facets of number comparison listed above can be addressed within this conceptual framework, surprisingly few attempts exist to account quantitatively for each of them simultaneously. The present study attempted to fill in this gap, and has led to several noteworthy conclusions.

First, random walk models provide a coherent conceptual framework to account for a considerable variety of aspects characterizing numerical comparison. This account is remarkably parsimonious, based on a minimal number of meaningful parameters, and relying mostly on the dynamic assumptions inherent to the basic processing model (Figure 1). Specifically, with only four individually interpretable parameters ($a, b, k,$ and M) the model accounts well for the mean correct latency and the error rate of all 72 individual digit pairs, and adding a fifth parameter (σ_M) provides a good account of the variability of the latencies, again on the single-pair level (Figure 3). In addition, we present strong evidence that the latency of incorrect responses also exhibits a numerical distance effect (Figure 5), as predicted by random walk models. Note that according to alternative processing models (e.g., fast-guess models, cf. Luce, 1986, ch. 7; for application in number comparison, see Dehaene, 1996, pp. 61ff.; Schwarz & Ischebeck, 2000) errors arise as a strategic ex ante selection – a risky bet – of one specific response to maximize speed, and are thus essentially unrelated to the digit pair actually presented (for a discussion of how neural network models might possibly account for errors, see Verguts

et al., 2005, pp. 77-78). It is noteworthy that in order to account for the effect of numerical distance on the latency of errors the random walk model relies exclusively on its structural assumptions, without adding any new assumptions, new mechanisms, or new parameters.

A second and related conclusion is that this remarkable ability to account quantitatively for many detailed aspects of number comparison is not simply a universal characteristic of the broad class of evidence accrual models (cf., Luce, 1986, ch. 8) in general, rather it requires more specific assumptions about the step size densities involved, and also about the mapping of numerical magnitude onto the drift rates driving the evidence accumulation process. All extant evidence accumulation models for paired numerical comparison designs assume that – in a stochastic sense – the accumulation process for, say, the digit pair $(8, 3)$ tends towards the upper response barrier in the same manner as it tends towards the lower response barrier for the complementary digit pair, $(3, 8)$. Within this widely accepted model frame, if the step size densities themselves are symmetric (e.g., normal; Poltrock, 1989; Sigman & Dehaene, 2005; Smith & Mewhort, 1998) then the model predicts that mean correct latency for $(8, 3)$ and mean error latency for $(3, 8)$ are necessarily the same, even when the response barriers are different. Our results (Figure 2) show conclusively that this fundamental prediction is systematically violated, so that the adequacy of the model shown in Figure 1 (assuming fixed parameters across trials) relies on assuming a skewed step size density. More specifically, the longer tail of this density must point towards the incorrect response barrier, leading to fast first passages across this boundary. Alternatively, mixture (e.g., fast-guess) models (Luce, 1986, ch. 7), or random walk, diffusion, or linear ballistic accumulator models assuming variability of, e.g., the starting point across trials (e.g., Brown & Heathcote, 2008; Donkin et al., 2011; Ester et al., 2014; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004;) are also able to predict faster error than correct responses by introducing additional free model parameters (e.g., s_Z and η in Ratcliff & Smith, 2004).

Similarly, random walk models of numerical comparison differ in their specific assumptions about how numerical magnitudes map onto those noisy internal representations whose aggregation defines the evidence accrual process. Therefore, they provide a well-defined conceptual framework to compare meaningfully the way in which numerical magnitudes are internally transformed and then aggregated. Our comparison of a linear vs. logarithmic number-to-drift mapping produced a much better account of our findings with the latter transformation, which is in line with several earlier related theoretical arguments and empirical findings (for detailed discussion, see Dehaene, 2003, 2011; Cordes et al., 2007; Nieder, 2005). Specifically, the logarithmic mapping accounts well for the Weber-law signature that we observed not only for mean SL, but also for hitherto untested aspects of numerical comparison such as the SD of SL and error rate (Figure 4).

Our participants responded by saccadic eye movements which in simple two-choice tasks are well-known to be faster and less variable than manual responses (e.g., Khalid and Ansorge, 2013, Figures 5 and 6). In addition, saccadic responses provide the opportunity to study not only latency but also other aspects of the response, such as, e.g., its amplitude. As a new form of a numerical distance effect we observed a systematically increasing saccadic response amplitude as the numerical distance between the two digits decreased (Figure 6); for a given distance this effect was more pronounced for larger magnitudes (Weber-law signature; bottom right panel of Figure 4). Can this new finding be interpreted in a coherent way within the present framework of random walk models?

One suggestion is provided by findings first reported by Mattes, Ulrich, and Miller (1997; for related findings, see Miller & Schröter, 2002). In their go-nogo RT study participants were provided with a visual advance cue indicating the go probability (0.1, 0.2, 0.4, 0.6, 0.8, or 1.0) for the current trial. As expected, mean RT decreased with the signaled go probability. However, Mattes et al. in addition measured and analyzed the peak force output of the manual responses (RF) and

observed that RF systematically increased with decreasing signaled go probability. That is, RF was highest when the cued go probability was lowest, so that the slowest responses were associated with the highest amplitude. Clearly, this pattern of an inverse relation between go probability (and thus speed) and the amplitude of manual responses observed by Mattes et al. corresponds quite closely to the inverse relation that we observed between numerical distance (and thus speed) and the amplitude of saccadic responses. A further parallel to our results is Mattes et al.'s finding that within any given cued go probability condition the correlation between speed and amplitude was close to zero, much as we found no difference in the saccadic amplitudes of slow vs. fast responses within a given numerical distance.

Relating the Mattes et al. (1997) account in terms of differential motor preparation to the conceptual framework of random walk models, one way to explain our findings, then, starts from noticing that when the numerical difference is small, the induced accumulation process is characterized by a small drift rate. Correspondingly, for digit pairs with a small numerical difference the process will on average spend a larger amount of time in close vicinity to the evidence boundary the crossing of which triggers the final motor command. Therefore, numerical distance and response amplitude will be inversely related if the motor command "primed" by a random walk closely approaching a response barrier is cumulatively prepared to a degree that increases with the duration ("dwell time") of walk positions that are already quite close to the boundary (for related models of covert motor preparation, see Servant, White, Montagnini, & Burle, 2015; Thura et al., 2014). As a simple illustration of our account in terms of cumulative covert motor preparation, for the best-fitting parameter values reported above we simulated the random walk model with double exponential step sizes, and compared the dwell time during which the random walk stayed less than c units away from the response barrier. For example, for the pairs $(1, 9)$ and $(8, 9)$ the mean "close-to-boundary" dwell times (using $c = 2$)

were, 1.0 and 4.2 ms respectively, with other values of c producing similar relations of dwell times. Considering the dwell times as a measure of cumulative motor preparation (leading to larger response amplitudes), these values are qualitatively consistent with our findings (Figure 6). Clearly, this tentative account requires more rigorous experimental tests; however, it does suggest a qualitative explanation that fits in well with the random walk concept of stochastic evidence accumulation.

We believe that our findings can have useful implications for studies in which number comparison represents a well-established diagnostic tool to address relevant open research topics which are currently actively investigated. For example, in a study designed to look at age effects on number comparison Norris et al. (2015) observed that older adults (mean age 65 years) were consistently slower than younger adults (mean age 20 years); at the same time, older adults were also considerably more accurate. An important question, then, is: how much slower would the older adults have been, had they operated at an accuracy level comparable to that of the younger adults? The random walk model presented here can profitably be exploited in these contexts, and is ideally suited to help address important questions of this type. Specifically, it would be highly informative to analyze on the basis of the model described here if the speed advantage observed for younger adults could be attributed to a large degree, or even completely, simply to the use of stricter (i.e., larger) response barriers with increasing age.

In conclusion, if combined with suitable specific assumptions (skew step size density, logarithmic number-to-drift mapping), random walk models provide a remarkably parsimonious (only four parameters) yet accurate account of many aspects involved in numerical comparison. In many research contexts using number comparison tasks, their meaningful, interpretable parameters offer a more coherent, detailed, and process-oriented interpretation of results than seems available on the basis of standard statistical analyses.

References

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, **19**, 967–983.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467–472.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, **57**, 153–178.

Buckley, P.B., & Gillman, C.B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, **103**, 1131–1136.

Cassey, P., Heathcote, A., & Brown, S.D. (2014). Brain and Behavior in Decision-Making. *PLoS Computational Biology*, **10**, e1003700.

Chen, Q., & Verguts, T. (2010). Beyond the mental number line: A neural network model of number-space interactions. *Cognitive Psychology*, **60**, 218–240.

Cohen Kadosh, R., Cohen Kadosh, K., Linden, D.E.J., Gevers, W., Berger, A., & Henik, A. (2007). The brain locus of interaction between number and size: A combined functional magnetic resonance imaging and event-related potential study. *Journal of Cognitive Neuroscience*, **19**, 957–970.

Cordes, S., Gallistel, C.R., Gelman, R., & Latham, P. (2007). Nonverbal arithmetic in humans: Light from noise. *Perception & Psychophysics*, **69**, 1185–1203.

Darling, D.A., & Siegert, A.J.F. (1953). The first passage problem for a continuous Markov process. *The Annals of Mathematical Statistics*, **24**, 624–639.

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

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, **8**, 4768.

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

Dehaene, S. (2011) *The number sense: How the mind creates mathematics, revised and updated edition*. New York, NY: Oxford University Press.

Dehaene, S., & E.M. Brannon, E.M. (Eds.) (2011). *Space, time and number in the brain*. London: Academic Press (Elsevier).

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, **18**, 61–69.

Ester, F.E., Ho, T.C., Brown, S.D., & Serences, J.T. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision*, **14**, 1–12.

Feller, W. (1971). *An introduction to probability theory and its applications*, vol. II. (2nd ed.) New York: Wiley.

Fernandez, S.R., Rahona, J.J., Hervas, G., Vazquez, C.,& Ulrich, R. (2011). Number magnitude determines gaze direction: Spatial-numerical association in a free-choice task. *Cortex*, **47**, 617–620.

Fischer, R., & Miller, J. (2008). Does the semantic activation of quantity representations influence motor parameters? *Experimental Brain Research*, **189**, 379–391.

Foley, J., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold vs. contrast (TvC) function for contrast discrimination. *Journal of the Optical Society of America, A*, **15**, 1036–1047.

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

Gallistel, C.R., & Gelman, R. (2005). Mathematical cognition. In Holyoak, K., & Morrison, R. (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 559-588). New York: Cambridge University Press.

Ganor-Stern, D., & Goldman, R. (2015). Reaching towards an end: Numerical end and distance effects in motor movements. *Journal of Cognitive Psychology*, **27**, 490–498.

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

Godwin, H., Hout, M.C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review*, **21**, 689–695.

Gold, J.I., & Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, **30**, 535–574.

Henik, A., & Tzelgov, J. (1982). Is 3 greater than 5 – the relation between physical and semantic size in comparison tasks. *Memory & Cognition*, **10**, 389–395.

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

Holloway, I.D., Battista, C., Vogel, S.E., & Ansari, D. (2013). Semantic and perceptual processing of number symbols: Evidence from a cross-linguistic fMRI adaptation study. *Journal of Cognitive Neuroscience*, **25**, 388–400.

Huber, S., Moeller, K., & Nuerk, H.-C. (2014). Dissociating number line estimations from underlying numerical representations. *The Quarterly Journal of Experimental Psychology*, **67**, 991–1003.

Hyde, D.C., Khanum, S., & Spelke, E.S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, **131**, 92–107.

Ischebeck, A. (2003). Differences between digit naming and number word reading in a flanker task. *Memory & Cognition*, **31**, 529–537.

Kamienkowski, J.E., Pashler, H., Dehaene, S., & Sigman, M. (2011). Effects of practice on task architecture: Combined evidence from interference experiments and random-walk models of decision making. *Cognition*, **119**, 81–95.

Khalid, S., & Ansorge, U. (2013). The Simon effect of spatial words in eye movements: Comparison of vertical and horizontal effects and of eye and finger responses. *Vision Research*, **86**, 6–14.

Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions: Theory and applications*. London: Imperial College Press.

Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology*, **103**, 546–565.

Link, S.W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, **12**, 114–135.

Link, S.W. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, **34**, 2–41.

Link, S.W. (1992). *The wave theory of difference and similarity*. Hillsdale, N.J.: L. Erlbaum Associates.

Livingstone, M.S., Srihasam, K., & Morocz, I.A. (2010). The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Animal Cognition*, **13**, 711–719.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, **1**, 476 - 490.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, **315**, 57–59.

Mattes, S., Ulrich, R., & Miller, J. (1997). Effects of response probability on response force in simple RT. *The Quarterly Journal of Experimental Psychology*, **50**, 405–420.

Miller, J., & Schröter, H. (2002). Online response preparation in a rapid serial visual search task. *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 1364–1390.

Milosavljevic, M., Madsen, E., Koch, C., & Rangel, A. (2011). Fast saccades toward numbers: Simple number comparisons can be made in as little as 230 ms. *Journal of Vision*, **11**, 1–12.

Moeller, K., Huber, S., Nuerk, H.-C., & Willmes, K. (2011). Two-digit number processing: holistic, decomposed or hybrid? A computational modelling approach. *Psychological Research*, **75**, 290–306.

Moore, A.M., & Ashcraft, M.H. (2015). Children's mathematical performance: Five cognitive tasks across five grades. *Journal of Experimental Child Psychology*, **135**, 1–24.

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

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, **115**, 10–25.

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, **6**, 177–190.

Norris, J.E., McGeown, W.J., Guerrini, C., & Castronovo, J. (2015). Aging and the number sense: preserved basic non-symbolic numerical processing and enhanced basic symbolic processing. *Frontiers in Psychology*, **6**, 999.

Pashler, H., & Johnston, J.C. (1998). Attentional limitations in dual-task performance. In Pashler, H.. (Ed.), *Attention* (pp. 155-189). Philadelphia: Taylor & Francis Press.

Poltrock, S.E. (1989). A random walk model of digit comparison. *Journal of Mathematical Psychology*, **33**, 131–162.

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, **90**, 1392–1407.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347–356.

Ratcliff, R., & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, **111**, 333–367.

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

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**, 1137–1147.

Schall, J.D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, **2**, 33–42.

Schwarz, W. (1990). Stochastic accumulation of information in discrete time: Comparing exact results and Wald–approximations. *Journal of Mathematical Psychology*, **34**, 229–236.

Schwarz, W. (1991). Variance results for random walk models of choice reaction time. *British Journal of Mathematical and Statistical Psychology*, **44**, 251–264.

Schwarz, W. (2001). The Ex–Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, and Computers*, **33**, 457–469.

Schwarz, W., & Eiselt, A.K. (2009). The perception of temporal order along the mental number line. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 989–1004.

Schwarz, W., & Eiselt, A.K. (2012). Numerical distance effects in visual search. *Attention, Perception & Psychophysics*, **74**, 1098–1103.

Schwarz, W., & H.–J. Heinze (1998). On the interaction of numerical and size information in digit comparison: A Behavioral and Event–Related Potential (ERP) Study. *Neuropsychologia*, **36**, 1167–1179.

Schwarz, W., & Ischebeck, A. (2000). Sequential effects in number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1606–1621.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number–size interference effects in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 507–522.

Schwarz, W., & Miller, J. (2014). When less equals more: Probability summation without sensitivity improvement. *Journal of Experimental Psychology: Human Perception and Performance*, **40**, 2091-2100.

Schwarz, W., & Stein, F. (1998). On the temporal dynamics of digit comparison processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 1275–1293.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, **48**, 630–633.

Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using covert response activation to test latent assumptions of formal decision-making models in humans. *The Journal of Neuroscience*, **35**, 10371–10385.

Shettleworth, S.J. (2010). *Cognition, evolution, and behavior*. (2nd ed.). Oxford: Oxford University Press.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, **3**, 334–349.

Smith, D.G., & Mewhort, D.J.K. (1998). The distribution of latencies constrains theories of decision time: A test of the random-walk model using numeric comparison. *Australian Journal of Psychology*, **50**, 149–156.

Smith, P.L. (1990). A note on the distribution of response times for a random walk with Gaussian increments. *Journal of Mathematical Psychology*, **34**, 445–459.

Soltész, F., Szücs, D., Dékány, J., Márkus, A., & Csépe, V. (2007). A combined event-related potential and neuropsychological investigation of developmental dyscalculia. *Neuroscience Letters*, **417**, 181–186.

Song, J.-H., & Nakayama, K. (2008). Numeric comparison in a visually-guided manual reaching task. *Cognition*, **106**, 994–1003.

Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 18116–18120.

Steinborn, M.B., Bratzke, D., Rolke, B., Gordijn, M.C.M., Beersma, D.G.M., & Ulrich, R. (2010). The effect of 40 hours of constant wakefulness on number comparison performance. *Chronobiology International*, **27**, 807–825.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, **21**, 259–263.

Thura, D., Cos, I., Trung, J., & Cisek, P. (2014). Context-dependent urgency influences speed–accuracy trade-offs in decision-making and movement execution. *The Journal of Neuroscience*, **34**, 16442–16454.

Townsend, J.T., & Ashby, F.G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

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

Verguts, T., & van Opstal, F. (2014). A delta-rule model of numerical and non-numerical order processing. *Journal of Experimental Psychology: Human Perception and Performance*, **40**, 1092–1102.

Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, **10**, 130–137.

Appendix: Basic model assumptions and properties

<u>Mirror-symmetric random walk models</u>

A basic symmetry assumption made by extant stochastic accumulation models of number comparison (e.g., Kamienkowski et al., 2011; Link 1990, 1992; Poltrock 1989; Schwarz, 2001; Schwarz & Stein, 1998, Sigman & Dehaene, 2005; Smith & Mewhort, 1998) is that the random walks induced by complementary digit pairs (i, j) and (j, i) are mirror-symmetric, that is, that their step sizes densities are related by

$$g_{ji}(x) \;=\; g_{ij}(-x) \tag{1}$$

Note that even if (1) holds $g_{ij}(x)$ and $g_{ji}(x)$ may *individually* still be asymmetrical.

One account leading to (1) is the differencing model which assumes a digit-specific density, say h_i (with mean μ_i), which describes the part of the step induced by digit i, and similarly a density h_j for the digit j. Loosely speaking, for the pair (i, j) the density h_i describes the (stochastic) pull of digit i towards "its" (i.e., upper) barrier a, and h_j the pull exerted by digit j towards "its" (i.e., lower) barrier $-b$; the overall step size is the resulting net force, i.e., the difference of these antagonistic forces. This gives for the step size density (e.g., Feller, 1971, ch. V)

$$g_{ij}(x) \;=\; \int_{-\infty}^{\infty} h_i(x + s) \cdot h_j(s)\, ds$$

which in turn implies (1). Several other processing assumptions equally lead to the basic mirror-symmetry representation (1) (see Link, 1975, 1992; Luce, 1986).

As is well known (e.g., Luce, 1986, ch. 8.3.3.; Schwarz, 1990, 1991; Townsend & Ashby, 1983, ch. 10), if in addition to condition (1) the (two-sided) Laplace transform of $g_{ij}(x)$ is symmetrical (as is, e.g., the case if the g_{ij} are normal distributions), then (1) implies that error latencies to (i, j) are on average equal to mean correct

RT to the complementary pair (j, i), even when $a \neq b$. Note that the numerical distance effect is typically computed as an average across complementary digit pairs, so that for a given distance mean correct and error RT are then predicted to be equally fast for random walk models obeying (1) if the step size densities $g_{ij}(x)$ have symmetrical Laplace transforms. On the other hand, if the condition (1) holds but the step size densities $g_{ij}(x)$ have asymmetrical Laplace transforms then mean correct RT for (i, j) and mean error RT for (j, i) will usually differ. In the next section, we describe one specific random walk model with this property. We emphasize that faster error than correct responses can also be predicted by mixture (e.g., fast-guess) models (Luce, 1986, ch. 7), or by random walk, diffusion, or linear ballistic accumulator models assuming variability of, e.g., the starting point across trials (e.g., Brown & Heathcote, 2008; Donkin et al., 2011; Ester et al., 2014; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004).

The double exponential random walk model

The double exponential model shown in Figure 1 assumes that for $i < j$ the step size density associated with the pair (i, j) has a double exponential distribution (Figures 1 and 7, see Kotz & Nadarajah, 2000; for background on this model in the context of perceptual decision making, see Foley & Schwarz, 1998; Schwarz & Miller, 2014), with associated cumulative distribution function

$$G_{ij}(x) \;=\; \exp\left[-e^{-(x+\gamma-\mu_{ij})}\right] \tag{2}$$

where μ_{ij} is the net drift rate of the induced walk, and $\gamma = 0.5772\ldots$ is the Euler-Mascheroni constant. That is, the mean of the step size random variable (rv) $\mathbf{X}_{ij}$ with the distribution function G_{ij} is equal to μ_{ij}, and its standard deviation is equal to $\pi/\sqrt{6} \approx 1.28$.

A characteristic feature of the double exponential step size density $g_{ij} = G'_{ij}$, shown in Figure 7 (left), is its positive skew, leading to the asymmetric shape (Fig-
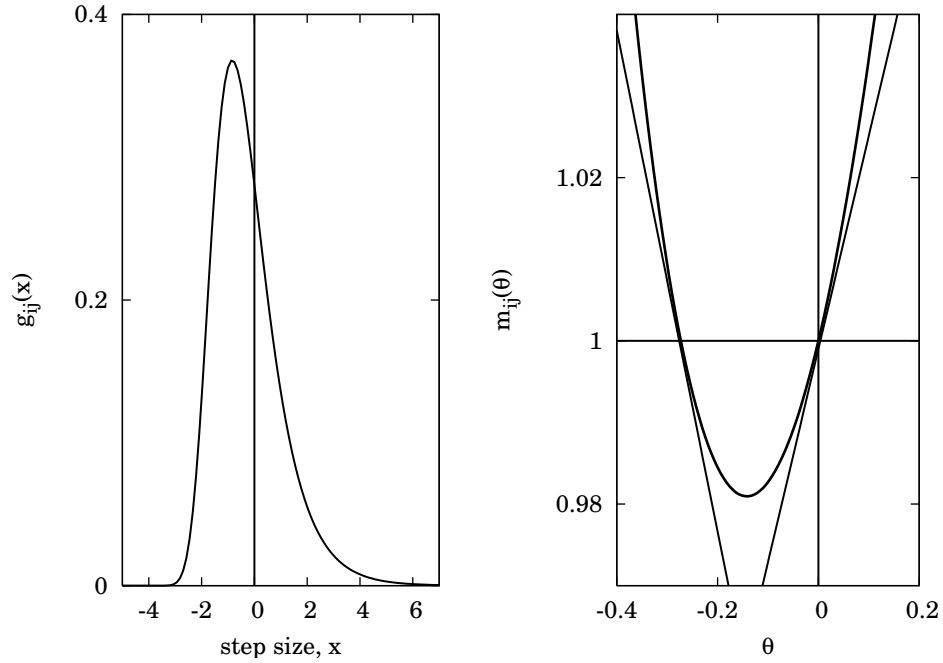
Figure 7. Left panel: skew double exponential step size density $g_{ij}(x)$, corresponding to the pair $(i = 6, j = 8)$, giving (using our estimate of $k = 0.907$) a mean step size equal to $\mu_{ij} = k \cdot \ln(i/j) = -0.261$. Right panel: The asymmetric two-sided Laplace transform $m_{ij}(\theta)$ of the step size density $g_{ij}(x)$, shown in the left panel. The two tangents to $m_{ij}(\theta)$ indicate the unequal (in absolute value) slopes at the two points where $m_{ij}(\theta) = 1$. This ratio of absolute slopes (here equal to 1.18) in turn determines the relation of mean correct RT for (i, j) to mean error RT for (j, i).

ure 7, right) of its two-sided Laplace transform,

$$m_{ij}(\theta) \;=\; \mathsf{E}[\exp(-\theta \mathbf{X}_{ij})] \;=\; \Gamma(1 + \theta) \cdot \exp[-\theta(\mu_{ij} - \gamma)] \tag{3}$$

where Γ is the Gamma function. As mentioned above, it is the asymmetry of m_{ij} which implies that mean correct RT for (i, j) will not be equal to mean error RT for (j, i). If the left digit is smaller than the right digit (i.e., $i < j$) then negative steps are directed towards the lower barrier at $-b$ associated with the correct response "right digit larger" (cf., Figure 1), and the random walk will on average tend to the lower response barrier, i.e., $\mu_{ij} < 0$. Specifically, in this case, the positive skew of $g_{ij}(x)$ implies that mean correct RT to (i, j) will be slower than mean error RT for (j, i). Conversely, the step size densities $g_{ji}(x)$ for the complementary digit pairs

$(j, i), j > i$ are given by the relation eq. (1), also leading on average to slower correct responses to (j, i) as compared to mean error RT for (i, j).

Number-to-drift mapping: Linear vs. logarithmic

The basic random walk model also provides a well-defined conceptual framework within which to compare meaningfully different assumptions about the mapping of numerical magnitude onto its internal representation. Specifically, we compared two conceptualizations of how numerical magnitude influences the net drift rate μ_{ij} of the double exponential random walk model: a linear and logarithmic model. In the linear model for each digit i

$$\mu_i = c + k \cdot i \tag{4}$$

whereas in the logarithmic model

$$\mu_i = c + k \cdot \ln(i) \tag{5}$$

Interpreting the component drift rate μ_i as the mean pull of the left digit i towards "its" barrier a, and similarly the component drift rate μ_j as the mean pull exerted by the right digit j towards "its" barrier $-b$, the resulting net drift μ_{ij} is the difference of these antagonistic mean forces, i.e.,

$$\mu_{ij} = \mu_i - \mu_j = \begin{cases} k \cdot (i - j) & \text{linear} \\ k \cdot [\ln(i) - \ln(j)] = k \cdot \ln\left(\frac{i}{j}\right) & \text{logarithmic} \end{cases} \tag{6}$$

i.e., the constant c cancels in both models (4, 5).

In effect, then, both the linear and the logarithmic model versions require a total of only four parameters to predict all (correct and error) mean RT and all error rates of 72 digit pairs: the scale parameter k, the response barriers a and $-b$, and the mean non-decisional latency component, μ_M. In addition, predictions of the

variance of the decision time require no new parameters at all because they follow from (2), and known general variance results concerning random walk models (see Schwarz, 1991); the only additional parameter required is the variance of the non-decisional latency component, σ_M^2.

<u>Definition of objective function</u>

The objective function f producing, for the model in Figure 1 with double exponential step sizes (eq. A.2) and logarithmic number-to-drift mapping (eq.s A.5/6), the fit shown in Figure 3 compares observed mean correct SLs and mean error rates of all 72 digit pairs (i, j) with the corresponding model prediction. Denote as SL_{ij}^o, SL_{ij}^p, respectively, the observed and predicted mean saccadic latencies to digit pair (i, j), where s.e.$_{ij}^{SL}$ is the estimated standard error of SL_{ij}^o. Similarly denote as ER_{ij}^o and ER_{ij}^p, respectively, the observed and predicted mean error rate, where s.e.$_{ij}^{ER}$ is the estimated standard error of ER_{ij}^o. The predicted values are a function of the four model parameters a, b, k, and μ_M. As explained in the text, the objective function was then

$$f(a, b, k, \mu_M) \;=\; \sum_{i=1}^{9} \sum_{j=1, j \neq i}^{9} \left[\left(\frac{SL_{ij}^o - SL_{ij}^p}{\text{s.e.}_{ij}^{SL}} \right)^2 + \left(\frac{ER_{ij}^o - ER_{ij}^p}{\text{s.e.}_{ij}^{ER}} \right)^2 \right]$$

Given the model has only four parameters, we evaluated and compared the points in the parameter space using a grid search with adjustable step sizes; this was then checked by Mathematica's (9.0.1, Wolfram Research) NMinimize function.

Chapter 3

Exploring the origin of the number size congruency effect: Sensitivity or response bias?

Abstract

Physical size modulates the efficiency of digit comparison, depending on whether the relation of numerical magnitude and physical size is congruent or incongruent (Besner & Coltheart, 1979), the number-size congruency effect (NSCE). In addition, Henik and Tzelgov (1982) first reported a NSCE for the reverse task of comparing the physical size of digits such that the numerical magnitude of digits modulated the time required to compare their physical sizes. Does the NSCE in physical comparisons simply reflect a number-mediated bias mechanism related to making decisions and selecting responses about the digit's sizes? Alternatively, or in addition, the NSCE might indicate a true increase in the ability to discriminate small and large font sizes when these sizes are congruent with the digit's symbolic numerical meaning, over and above response bias effects. We present a new research design that permits us to apply signal detection theory to a task that required observers to judge the physical size of digits. Our results clearly demonstrate that the NSCE cannot be reduced to mere response bias effects, and that genuine sensitivity gains for congruent number-size pairings contribute to the NSCE.

Introduction

Numbers and numerical information play a central role in how humans represent, communicate, and respond to quantity-related aspects of their environments. Correspondingly, a considerable amount of research has addressed many cognitive, neuronal, developmental, and clinical aspects of how we process quantitative information (for general review, see Dehaene, 2011; Nieder, 2005).

According to one influential model of number comparison (Moyer & Landauer, 1967; for recent formulations and review, see Ditz & Nieder, 2016; Ganor-Stern & Goldman, 2015; Gilmore, Attridge, & Inglis, 2011; Inglis & Gilmore, 2014; Maloney, Risko, Presto, et al., 2010; Nuerk, Moeller, Klein, et al., 2011; Reike & Schwarz, 2016; Sigman & Dehaene, 2005), digits are automatically converted into percept-like analog representations that are then in turn compared with each other, much like sensory representations of physical attributes such as brightness or orientation. One line of evidence consistent with this model derives from conflict paradigms in which the to-be-compared digits are presented with varying physical (i.e., font) sizes. According to the analog representation model, the digit's (task-irrelevant) physical size should modulate the efficiency of the comparison process, depending on whether the relation of numerical magnitude and physical size is congruent (as in $8 - 2$) or incongruent (e.g., $8 - 2$). This number-size congruency effect (NSCE) has indeed first been reported by Besner and Coltheart (1979), and has since then often been confirmed and extended (e.g., Algom, Dekel & Pansky, 1996; Cohen Kadosh, Cohen Kadosh, Linden, et al., 2007; Fitousi, 2014; Girelli, Lucangeli, & Butterworth, 2000; Henik & Tzelgov, 1982; Kaufmann, Koppelstaetter, Delazer, et al., 2005; Pansky & Algom, 1999; Schwarz & Heinze, 1998; Schwarz & Ischebeck, 2003; Szűcs & Soltész, 2007; Takahashi & Green, 1983; Tzelgov, Meyer, & Henik, 1992). Henik and Tzelgov (1982) first studied the reverse task of comparing the physical size of digits, independent of their numerical meaning.

Contrary to strictly serial models in which the physical features of the digit (including its size) are fully identified before the digit's numerical magnitude is accessed they observed reliable NSCEs such that the (logically task-irrelevant) numerical value of the digits modulated the time required to compare their physical sizes (for recent related studies, see Arend & Henik, 2015; Cantlon, Platt, & Branno, 2009; Faulkenberry, Cruise, Lavro, & Shaki, 2016; Gabay, Leibovich, Henik, & Gronau, 2013; Goldfarb & Treisman, 2010; Pina, Castillo, Cohen Kadosh, & Fuentes, 2015; Reber, Christensen, & Meier, 2014; Risko, Maloney, & Fugelsang, 2013; Santens & Verguts, 2011; Sobel, Puri, & Faulkenberry, 2016).

An as-yet-open question, addressed in the present experiment, is that the NSCE in physical comparisons might simply reflect a number-mediated bias mechanism related to making decisions and selecting responses about the digit's sizes (e.g., Risko et al., 2013; Schwarz & Heinze, 1998; Sobel et al., 2016). Expressed in the terminology of signal detection theory (SDT; Macmillan & Creelman, 2005) one interpretation of the NSCE is that the digit's numerical magnitude may induce a number-related response bias because observers implicitly tend to associate numerical magnitude with physical size. In this view, congruent pairings of physical size and numerical magnitude do not actually enhance the sensitivity of the size discrimination process but more simply benefit from a tendency to respond more readily "larger" ("smaller") with numerically large (small) digits. Alternatively, the NSCE might more adequately be interpreted as reflecting a genuine sensitivity benefit tied to the numerical meaning of the digits presented (for example, Sobel, Puri, Faulkenberry, & Dague, 2017, Exp. 3 report that numerical information can influence but probably not guide visual search). In this interpretation, the NSCE reflects a true increase in the ability to discriminate small and large font sizes when these sizes are congruent with the digit's symbolic numerical meaning – over and above mere response tendencies.

With respect to these alternatives, it is unfortunate that most extant research designs exploring the NSCE are based on the analysis of response time. Chronometric measures often help to investigate the onset and time course of the NSCE but they also tend to make it difficult to distinguish between bias and sensitivity accounts (e.g., Luce, 1986, ch.s 3-4; Matt & Dzhafarov, 2014). Therefore, in the present study we developed a new research design to apply SDT to a task in which observers judged the physical size of digits. Specifically, single digits were presented in either of two global sizes (sets S(mall) and L(arge)); *between* these sets digits differed massively in size. In addition, *within* each size set digit exemplars could have a slightly smaller ($S-$, $L-$) or larger ($S+$, $L+$) font size (for details, see Methods and Figure 1), and the task of the observer was to classify the digit shown as being a smaller or a larger exemplar within its size set. In this design, the difference between "global" font size (S vs. L) is large (and was made even more obvious by using different colors to indicate the different size sets) but not task-relevant, whereas the task-relevant feature (the "local" size within each set) requires a more difficult discrimination of minute size differences.
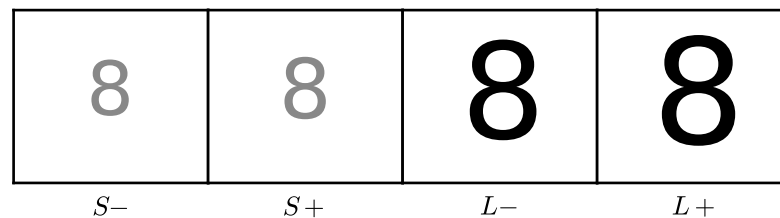

Figure 1. Sizes of digit stimuli as used in the present experiment. For example, the digit 8 was shown in four different physical sizes. The two left panels illustrate the small size set (S); the two right panels illustrate the large size set (L). Both size sets consist of two slightly different sizes: smaller ($-$) and larger ($+$) exemplars. Digits from one size set were presented in dark blue and from the other size set in brown. The two left digits ($S-$, $S+$) represent incongruent conditions because the global physical size (small, S) does not correspond to the numerical magnitude (large, >5). Similarly, the two right digits ($L-$, $L+$) represent congruent conditions because the global physical size (large, L) corresponds to the numerical magnitude (large, >5).

The aim of the present study was to analyze results from this size identification design within the framework of SDT. In particular, we aimed to extract, and to compare, separate bias and sensitivity measures for congruent and incongruent pairings of numerical magnitude and (global) physical size to further clarify the origin and nature of the NSCE.

Method

Participants. 36 University of Potsdam students, aged 18-39, with normal or corrected-to-normal vision participated in one session of approximately one hour. They received a payment of €8 or course credit for their participation.

Stimuli and apparatus. The stimuli consisted of the digits 1 – 9, excluding 5. Each trial display consisted of a single digit presented against a gray background in Verdana font on a 75-Hz, 1028 × 768 px VGA color monitor. As shown in Figure 1, we divided the stimuli into two sets: a small size set (S), and a large size set (L). The mean font size for the small size set was 30 pixel (px; 1 deg = 65 px at the viewing distance of approximately 114 cm); for the large size set it was 52.5 px. Therefore, the mean size distance of the fonts of these sets was 22.5 px. Within both size sets the digits were presented in two slightly different physical sizes: smaller ($-$) and larger ($+$) exemplars. Within the small size set the digits were presented in a font size of 28 px ($S-$) or 32 px ($S+$), whereas within the large size set they were presented with font sizes of 50 px ($L-$) or 55 px ($L+$), making the within-set size differences (S: 4 px; L: 5 px) much smaller than the mean between-set difference (22.5 px). In addition, the digits from one size set were presented in dark blue and from the other size set in brown (counterbalanced between subjects) in order to avoid any uncertainty about the size set (S or L) to which the digit presented belonged (as variations of the color-to-size mapping had no significant main or interaction effect, this factor is ignored in the sequel).

Procedure. In each of ten blocks, all eight digits were presented twice in each of the four sizes in random order. Each trial started with the presentation of a fixation cross; after a delay of 500 ms the digit was presented until a response was given. The task was to indicate by button press whether the presented digit was a physically smaller ($-$) or a larger ($+$) exemplar within its size set. Following the response, the participants received visual feedback about the correct answer and the response they had given. A blank screen was presented after 1000 ms for 500 ms, before the next trial started. After each block, the overall error rate was computed for this block separately for each size set (S and L). In the first block, the error rate had to be between 15 per cent and 30 per cent, in block two until block four between 10 per cent and 35 per cent, and after the fourth block between 5 per cent and 40 per cent. If the error rate of one size set did not meet these criteria then the font size of the larger ($+$) exemplars of this set was changed one px in the required direction. For example, if the error rate for the digits in the small size set (S) in the first block was below 15 per cent then the size of the larger exemplars of the small digits ($S+$) was reduced from 32 px to 31 px. Required size changes were applied after each block.

Preliminary Data Reduction. Only blocks in which the error rates for both size sets were within the targeted accuracy range were included in the analyses. The first block was considered practice and was always excluded. According to the block inclusion rules described, on average 5.5 (SD=1.4) blocks for each partici-pant were included in the analysis; including all blocks, except block 1, produced the same pattern of in-/significant results. In the present design, a digit presented in any trial is characterized by three independent attributes: its membership in the physical size set L or S, its numerical magnitude (<5 vs. >5), and by being a smaller ($S-$, $L-$) or larger ($L+$, $S+$) exemplar within its size set. Only the last attribute is response relevant; the first two attributes are logically and statistically independent of each other, and of the correct response. Congruency was defined

by the correspondence of the two response-irrelevant attributes physical size and numerical magnitude, so that digits <5 in the set S and digits >5 in the set L were congruent, whereas digits >5 in the set S and digits <5 in the set L were incongruent. Thus, for each participant the sensitivity index d' and the response bias index $\ln \beta$ (Macmillan & Creelman, 2005) were computed separately for two × two conditions, physically small or large (defined by set S vs. L) × numerically small or large (defined by magnitude <5 vs. >5).

Results

Sensitivity index d'. Values of d' for each participant were subjected to a 2-factorial ANOVA with the two within-subject factors physical size set (2: S/L) and numerical magnitude (2: $<5/>5$).

As intended, the required discrimination of sizes within each set was not an easy task, yielding an overall d' of 1.42 (s.e. 0.09) (corresponding to 74.4 % correct responses, relative to a chance level of 50 %).

Neither physical size set nor numerical magnitude had a main effect on d'. Central to our study, the interaction between physical size set and numerical magnitude was significant (Figure 2, left panel), F(1, 35) = 9.82, MSE = 0.15, p $<$ .01, $\eta^2 =$ .22. Differences in d' were found within each size set separately: for physically small digits the difference between numerically small ($d' = 1.54$) and numerically large digits ($d' = 1.40$) was significant, t(35) = -1.78, p $<$ .05, η^2 = .18. Similarly, for physically large digits the d' difference between numerically small digits ($d' = 1.24$) and numerically large digits ($d' = 1.50$) was significant, t(35) = 3.06, p $<$ .01, η^2 = .43. Thus, the congruency effect found did not depend on just one specific physical size set, as shown in Figure 2.

In order to validate and further investigate these findings we also considered the congruency effect (defined as the d' difference between congruent and incongruent

conditions) for each participant separately. On average, d' was 0.21 (s.e. 0.07) larger for congruent than for incongruent conditions; the participants answered correctly in 76.1 per cent of all congruent, and in 72.8 per cent of all incongruent trials. Overall, 25 of 36 participants (69%) showed a positive difference between the d's of congruent vs. incongruent conditions, with the mean positive differences (0.40) being larger than the mean negative differences (-0.23), W+ = 515, W− = 151, p < .01, η^2 = .11, Wilcoxon signed-rank test.

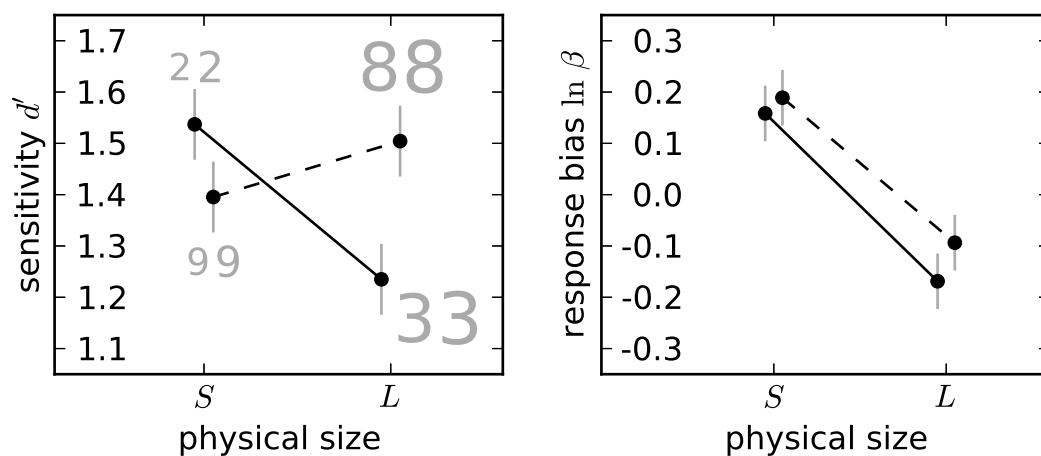**Figure 2.** Sensitivity index d' and response bias index $\ln \beta$ for each combination of physical size (set S vs. L) and numerical magnitude (<5 vs. >5). Left panel: Circles show the mean sensitivity index d' (ordinate) for numerically small (<5; solid line) and large (>5; dashed line) digits as a function of size set ((S)mall / (L)arge; abscissa). Each condition is illustrated by one smaller ($-$) and one larger ($+$) digit exemplar. Right panel: Circles show the mean response bias index $\ln \beta$ (ordinate) for numerically small (solid line) and large (dashed line) digits as a function of size set (abscissa). Error bars represent ± 1 SE (Loftus & Masson, 1994).

Response bias index $\ln \beta$. Values of $\ln \beta$ (positive values indicating a tendency to respond "smaller") for each participant were subjected to a repeated-measures ANOVA with the same factors as used for d'.

As expected under our balanced design, the grand mean of $\ln \beta$ was close to zero (0.02; s.e. 0.09), indicating that our participants had no overall bias towards using one of the two responses. Only the factor size set exerted a systematic main

effect on $\ln \beta$. Participants tended to classify the digits as being smaller ($-$) if they belonged to the small size set (S), and as being larger ($+$) if they belonged to the large size set (L), F(1, 35) = 7.83, MSE = .10, p < .01, η^2 = .18. As shown in Figure 2 (right panel), the mean response bias index was $+0.17$ for digits belonging to size set S, and -0.13 for digits belonging to size set L. Numerical magnitude had no main effect on $\ln \beta$. Finally, differences in $\ln \beta$ between congruent and incongruent trials were quite small in both size sets, as reflected in the nonsignificant interaction of physical size and numerical magnitude, F(1, 35) = .18, MSE = .10, p = .67, η^2 = .01.

Discussion

As reviewed in the Introduction, judgments about the physical size of digits are faster and less error-prone when physical size and numerical magnitude are congruent rather than incongruent (Henik & Tzelgov, 1982; for recent review, see Arend & Henik, 2015; Fitousi, 2014). Does this well-established performance benefit reflect a genuine enhancement of sensitivity, that is, an increase in the ability to discriminate small and large font sizes when numerical magnitude and physical size are congruent? Or is it more adequately attributed to a number-mediated response bias mechanism, for example, a tendency to respond more readily "larger" with numerically large, and more readily "smaller" with numerically small digits?

The present results strongly suggest that, in a general sense, response biases indeed systematically influence the way in which observers judge the physical size of digits. Specifically, as shown in Figure 2 (right panel), digits in the size set S are more readily classified as "smaller", and digits in the size set L more readily as "larger", even though smaller ($S-$, $L-$) and larger ($S+$, $L+$) exemplars were equally frequent in both size sets. If such biases selectively favor congruent number-size pairings and if "more readily" in the SDT sense translates into faster

responses (e.g., Luce, 1986, ch. 7), then the standard finding of shorter response times for congruent number-size pairings could be attributed to similar response bias mechanisms.

However, our results clearly demonstrate that the NSCE cannot simply be reduced to bias effects, and that genuine sensitivity gains for congruent number-size pairings contribute to the NSCE over and above mere response tendencies[1]. Specifically, as shown in Figure 2 (left panel), $S-$ vs. $S+$ digits in the size set S are discriminated with higher sensitivity when they are numerically small than when they are numerically large. Conversely, $L-$ vs. $L+$ digits in the size set L are discriminated with higher sensitivity when they are numerically large than when they are numerically small. Note that in our design within either size set S and L separately, each numerical magnitude was presented as a smaller and larger exemplar equally often; therefore, the differential sensitivity effects obtained cannot be attributed to overall response biases. Consider, for example, the simple biased response strategy of classifying numerically small digits more readily as "smaller", and numerically large digits more readily as "larger". In our design, this biased strategy would not produce the differential sensitivity effect shown in the left panel of Figure 2 because numerically small and large digits were presented equally often as small and large exemplars in both size sets. Also, our results provide no evidence for this specific form of number-related response bias (Figure 2, right panel). As discussed above, a related type of simple biased response strategy is to classify digits in size set S more readily as "smaller", and digits in size set L more readily as "larger". Again, this biased strategy (which is more prominent in our data; see Figure 2, right panel) would not produce the differential sensitivity effect observed because smaller and larger exemplars were presented equally often within both size sets. The same conclusion applies to any biased response

---

[1]Note that in the context of signal detection theory partitioning the total congruency effect into an inhibitory and a facilitatory component (as in RT studies involving a neutral condition; Henik & Tzelgov, 1982) requires a more elaborate SDT design (MacMillan & Creelman, 2005, ch. 5).

strategy based on some form of combination of numerical magnitude (<5 vs. >5) and physical size (S vs. L).

The present results suggest that the to-be-judged physical sizes of the digits ($S-$ vs. $S+$, and $L-$ vs. $L+$) are internally represented in a noisy format that is modifiable by numerical magnitude. This claim is consistent with the general theory of magnitude (ATOM) positing that there is a common cortical metric underlying several quantity-related attributes, such as space, time, and number (e.g., Bonato, Zorzi, & Umiltà, 2012; Cohen Kadosh, Lammertyn, & Izard, 2008; Eiselt & Nieder, 2012; Henik, Leibovich, Naparstek, et al., 2012; Leon & Shadlen, 2003; Schwarz & Eiselt, 2009; Walsh, 2003; Winter, Marghetis, & Matlock, 2015). For example, in their coalescence diffusion model Schwarz and Ischebeck (2003) proposed that information from task-irrelevant attributes (e.g., numerical magnitude, and physical font size S vs. L) often cannot be completely ignored. In their model, task-irrelevant information effectively combines with (and thereby modifies) information from task-relevant stimulus attributes to form an amalgam representation reflecting both, relevant and irrelevant stimulus aspects. On the face of it, our findings could be seen as being more compatible with an interaction at an early representational rather than at a late decision stage (e.g., Risko et al., 2013; Schwarz & Heinze, 1998; Sobel et al., 2016; Szűcs & Soltész, 2007); however, it should be stressed that SDT models per se are mute with respect to chronometric aspects. More generally, our findings clearly fit in with, and further extend, previous results from a variety of perceptual tasks (e.g., Casarotti, Michielin, Zorzi, & Umiltà, 2007; Corbett, Oriet, & Rensink, 2006; Fischer, Castel, Dodd, & Pratt, 2003; Godwin, Hout, & Menneer, 2015; Nieder, 2005; Schwarz & Eiselt, 2009, 2012; Sobel et al., 2016) suggesting that symbolic numerical meaning is extracted at an early processing stage from visual displays containing digits, and under favorable (i.e., congruent) conditions, may enhance perceptual sensitivity, even in basic psychophysical tasks involving judgments about physical size.

References

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition*, **24**, 557–572.

Arend, I., & Henik, A. (2015). Choosing the larger versus choosing the smaller: Asymmetries in the size congruity effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 1821–1830.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467–472.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, **36**, 2257–2273.

Cantlon, J.F., Platt, M.L., & Branno, E.M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, **13**, 83–91.

Casarotti, M., Michielin, M., Zorzi, M., & Umiltà, C. (2007). Temporal order judgment reveals how number magnitude affects visuospatial attention. *Cognition*, **102**, 101–117.

Cohen Kadosh, R., Cohen Kadosh, K., Linden, D.E.J., Gevers, W., Berger, A., & Henik, A. (2007). The brain locus of interaction between number and size: A combined functional magnetic resonance imaging and event-related potential study. *Journal of Cognitive Neuroscience*, **19**, 957–970.

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

Corbett, J.E., Oriet, C., & Rensink, R.A. (2006). The rapid extraction of numeric meaning. *Vision Research*, **46**, 1559–1573.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics, revised and updated edition*. New York, NY: Oxford University Press.

Ditz, H.M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of Royal Society B*, **283**, 20160083.

Eiselt, A.–K., & Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *The Journal of Neuroscience*, **33**, 7526–7534.

Faulkenberry, T.J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, **163**, 114–123.

Fischer, M.H., Castel, A.D., Dodd, M.D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, **6**, 555–556.

Fitousi, D. (2014). On the internal representation of numerical magnitude and physical size. *Experimental Psychology*, **61**, 149–163.

Gabay, S., Leibovich, T., Henik, A., & Gronau, N. (2013). Size before numbers: Conceptual size primes numerical value. *Cognition*, **129**, 18–23.

Ganor-Stern, D., & Goldman, R. (2015). Reaching towards an end: Numerical end and distance effects in motor movements. *Journal of Cognitive Psychology*, **27**, 490–498.

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology*, **64**, 2099–2109.

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

Godwin, H.J., Hout, M.C. & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review*, **21**, 689–695.

Goldfarb, L., & Treisman, A. (2010). Are some features easier to bind than others? The Congruency Effect. *Psychological Science*, **21**, 676–681.

Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., & Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Frontiers in Human Neuroscience*, **5**, 186.

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

Inglis, M., & Gilmore, J. (2014). Indexing the approximate number system. *Acta Psychologica*, **145**, 147–155.

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., Felber, S., & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *NeuroImage*, **25**, 888–898.

Leon, M.I., & Shadlen, M.N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*, **38**, 317–327.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, **1**, 476 - 490.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Macmillan, N.A., & Creelman C.D. (2005). *Detection theory: A user's guide (2nd ed.)*. Mahwah, NJ: Erlbaum.

Maloney, E.A., Risko, E.F., Preston, F., Ansari, D., & Fugelsang J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, **134**, 154–161.

Matt, J., & Dzhafarov, E.N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, **121**, 1–32.

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

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, **6**, 177–190.

Nuerk, H.–C., Moeller, K., Klein, E., Willmes, K., & Fischer, M.H. (2011). Extending the mental number line a review of multi-digit number processing. *Journal of Psychology*, **219**, 3–22.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 39–58.

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.

Reber, R., Christensen, B.T., & Meier, B. (2014). Effects of meaning and symmetry on judgments of sizes. *Frontiers in Psychology*, **5**, 1270.

Reike, D., & Schwarz, W. (2016). One model fits all: Explaining many aspects of number comparison within a single coherent model–A random walk account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. ,**42**, 1957-1971.

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**, 1137–1147.

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Journal of Experimental Psychology: Human Perception and Cognition*, **118**, 94–110.

Schwarz, W., & Eiselt, A.–K. (2009). The perception of temporal order along the mental number line. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 989–1004.

Schwarz, W., & Eiselt, A.–K. (2012). Numerical distance effects in visual search. *Attention, Perception & Psychophysics*, **74**, 1098–1103.

Schwarz, W., & Heinze, H.–J. (1998). On the interaction of numerical and size information in digit comparison: A behavioral and event–related potential (ERP) study. *Neuropsychologia*, **36**, 1167–1179.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number–size interference effects in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 507–522.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, **3**, 334–349.

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**, 1324–1336.

Sobel, K.V., Puri, A.M., Faulkenberry, T.J., & Dague, T.D. (2017). Visual search for conjunctions of physical and numerical size shows that they are processed

independently. *Journal of Experimental Psychology: Human Perception & Perfor-mance*, **43**, 444–453.

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia*, **45**, 3190–3202.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, **21**, 259–263.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 166–179.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, **7**, 483–488.

Winter, B., Marghetis, T., & Matlock, T. (2015). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, **64**, 209–224.

Chapter 4

Local probability effects of repeating irrelevant attributes

Abstract

We present two experiments in which participants classify stimuli having two potentially conflicting attributes, one of which is response-relevant whereas the other ("irrelevant") attribute is logically and statistically independent of the response. We introduce a novel design not used with filtering tasks before in which the main factor is the local (i.e., one-step) transition probability π (= 0.25, 0.50, 0.75) that the irrelevant attribute is repeated from one trial to the next. Experiment 1 involved a visual Simon task in which the color of the stimulus is relevant and its location is irrelevant. Experiment 2 used a semantic classification task in which the parity of a digit presented is relevant and its numerical magnitude is irrelevant. The results of both experiments demonstrate that participants in the π = 0.75 group responded faster when the irrelevant attribute is in fact repeated rather than alternated; in contrast, participants in the π = 0.25 group responded faster (Experiment 1) or equally fast (Experiment 2) when the irrelevant attribute is alternated rather than repeated. These expectancy-related effects cannot be attributed to spurious design contingencies as the irrelevant attribute was independent of the relevant attribute (and thus of the response), of the congruency status, and also of their alternation/repetition. One interpretation of our findings is that information about the irrelevant attribute in the previous trial is used much as an informative central precue, so that participants can prepare early processing stages in the current trial, with the corresponding benefits and costs typical of standard cueing studies.

Introduction

Among the most widely used experimental designs in human information processing are filtering tasks in which the stimuli presented to the observers for classification vary on two dimensions, or attributes, one of which is response-relevant, and one is not (e.g., Kahneman & Treisman, 1984; Pashler, 1998). Often, these attributes are independent, both logically and statistically, but are related by other, potentially conflicting aspects, in which case the generic term "conflict task" is sometimes used (e.g., Jimenéz & Méndez, 2013; Miller 1991; Schwarz & Ischebeck, 2003). These aspects may be defined by the relation of the irrelevant attribute to the relevant attribute (e.g., Stroop, 1935; Eriksen & Eriksen, 1974) or to the responses (e.g., spatial correspondence; Simon & Berbaum, 1990). Thus, standard designs such as the Stroop, the Simon, or the flanker task are all specific cases of conflict tasks, although this formal analogy should not conceal the several systematic differences that exist between them (for detailed review, see, e.g., Kornblum & Stevens, 2002; Pratte, Rouder, Morey, & Feng, 2010). In many conditions typical of conflict tasks the irrelevant stimulus attribute, even if uninformative about the relevant attribute and the response, nevertheless systematically influences the participant's performance, such that congruent irrelevant attributes lead to faster and more accurate responses than incongruent ones. Thus, speaking of attributes as "irrelevant" does not imply that they do not influence the way in which a stimulus is processed – it merely indicates that from knowledge about these attributes alone the correct response cannot be deduced at a better than chance level.

Effects of repeating irrelevant attributes

Early accounts of filtering and conflict tasks tended to focus on factors related to the specific conditions prevailing within any given trial (for reviews, see, e.g., Lu & Proctor, 1995; MacLeod, 1991; Pashler, 1998). However, subsequent research has documented systematic sequential effects in many of these tasks. An

extensive literature addresses sequential effects in conflict paradigms defined by the congruency status (i.e., congruent vs. incongruent) of the previous trial; a basic finding is that congruency effects tend to be reduced following incongruent as compared to congruent trials (e.g., Botvinick, Braver, Barch, Carter,& Cohen, 2001; Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014; Gratton, Coles, & Donchin, 1992; Jimenéz & Méndez, 2013; Mayr, Awh, & Laurey, 2003; Pfister, Schröder, & Kunde, 2013; Puccioni & Vallesi 2012).

In contrast, the present study addresses effects related to whether the irrelevant attribute of the previous trial is repeated or alternated. Extant findings regarding these repetition effects present a mixed picture, with several differences between paradigms and conditions. For example, using a standard Simon task Notebaert, Soetens, and Melis (2001; for related studies and reviews, see, e.g., Hommel, 2011; Notebaert & Soetens, 2003; Wühr & Ansorge, 2005; Zeischka, Deroost, Maetens, & Soetens, 2010) studied the effect of repeating the color (relevant attribute) and/or the location (irrelevant attribute) of the visual stimulus. They found (Experiments 1 and 2) that repeating the location had no main effect on RT but interacted with repeating the relevant attribute (and thus the response) such that RT was shorter when both attributes were repeated or both alternated (for similar findings, see Kleinsorge, 1999; Schwarz & Ischebeck, 2000). Furthermore, for response-to-signal intervals (RSIs) typical of standard conflict paradigms (1000 ms) repeating vs. alternating the irrelevant location did not influence the size of the Simon effect, whereas for very short RSIs (50 ms) a Simon effect was observed for location alternations but not for repetitions.

Related results on repeating irrelevant attributes were obtained by Kornblum and Stevens (2002, Experiment 4) in a four-choice task in which central letters (relevant attribute) were flanked either by response-incongruent arrows, or by incongruent letters that were targets on other trials. Repeating irrelevant arrow flankers had neither a main nor any interaction effect. However, repeating irrelevant flanking

letters reduced RT by 26 ms relative to alternating them, but even then there was no interaction between repeating the relevant and repeating the irrelevant attribute, in contrast to the study of Notebaert et al. (2001) using the Simon task.

Finally, effects of repeating irrelevant attributes have also been observed in non-spatial filtering tasks (Frings, Rothermund, & Wentura, 2007; Rothermund, Wentura, & De Houwer, 2005), for example, in semantic classification, such as numerical comparison (Dehaene, 1996; Pashler & Baylis, 1991; Pfister et al., 2013; Schwarz & Ischebeck, 2000; Tan & Dixon, 2011). For example, in Schwarz and Ischebeck (2000) participants classified numbers (presented as digits or number words; the irrelevant attribute) as smaller or larger than 5 (relevant attribute). Although uninformative about the response, repeating the logically irrelevant notation (arabic vs. verbal) produced a partial repetition benefit, even when the number itself was not repeated (for related partial repetition effects in digit naming, see Marcel & Forrin, 1974; Meuter & Allport, 1999).

Local repetition probability

The present study further investigates these sequential effects of repeating irrelevant stimulus attributes reviewed above by a technique that has not previously been used in the context of filtering and conflict tasks, namely by systematically varying the local, i.e., trial-to-trial repetition probability of the response-irrelevant attribute. In a classical paper, Bertelson (1961) first showed that local stimulus transition probabilities may be varied without altering the global (i.e., overall) probabilities of the stimulus alternatives involved. Specifically, he considered two-choice designs in which according to a first-order two-state Markov chain in each trial the local probability of repeating the previous stimulus was equal to $0 < \pi < 1$ while the global probability of each stimulus was held constant at 0.50. As seems natural, Bertelson (1961) used stimuli that varied only in one (thus, response-relevant) attribute, namely, the location at which they were presented. For RSIs typical of conflict tasks (500 ms), he found (his Table 4, group TL) that relative to a control

group ($\pi = 0.50$) observers in a $\pi = 0.75$ group responded faster to stimulus repetitions and slower to stimulus alternations, whereas an exactly complementary pattern was observed in a $\pi = 0.25$ group. Bertelson (1961) concluded that with sufficiently long RSIs his observers selectively prepared for the locally more likely stimulus alternative, even when for all three groups both stimuli were presented equally often. Subsequent research has confirmed and extended these basic findings, for example, to designs with more than two stimuli (Kornblum, 1968), to sequential trial dependencies of higher order (e.g., Jentzsch & Leuthold, 2005), and to analyses of event-related brain potentials (e.g., Leuthold & Sommer, 1993); for detailed reviews, see, for example, Jimenéz and Méndez (2013), Jones, Curran, Mozer, & Wilder (2015), or Soetens (1998). Notwithstanding these confirmations and extensions, a central aspect of these studies in the present context remains that all of them focused on the local repetition probability of the attribute that was actually response-relevant.

The main question, then, addressed by the present experiments is: Do Bertelson's (1961) conclusions regarding the local repetition probability of response-relevant attributes extend to response-irrelevant attributes as well? Specifically, adaptations to local repetition probabilities are often thought to reflect expectancies (for recent reviews, see Jimenéz & Méndez, 2013; Jones et al., 2015) enabling the participants to selectively prepare for more likely decisions or responses. Given that by definition irrelevant attributes per se provide no basis to anticipate the forthcoming response, one plausible view is that variations of the local repetition probability of irrelevant attributes will have no additional differential effect, over and above the well-documented basic effects of repeating irrelevant attributes reviewed above, all of which were obtained for $\pi = 0.50$. According to this view, repeating, for example, the irrelevant location in a Simon task would still show the effects (e.g., Notebaert et al., 2001) reviewed above, but these effects would not in turn be further modified by varying the local probability π of repeating the irrel-

evant location. Consistent with this view, note that repeating the relevant attribute necessarily implies repeating the response as well; in contrast, repetitions of the irrelevant attribute are by definition unrelated to the response. Therefore, to the degree that the selective preparatory effects of local probability first observed for relevant attributes by Bertelson (1961) mainly reflect response-related expectancies it is quite conceivable that these effects will not extend to variations of the local repetition probability π of attributes which are response-irrelevant.

On the other hand, at least some evidence suggests that variations of the local repetition probability of response-relevant attributes might have some effect. First, an extensive literature, referred to above, has clearly established that repeating irrelevant attributes per se does have reliable and systematic effects, under a wide range of well-understood conditions. Therefore, selectively adapting to the local probability of such repetitions might potentially enhance the efficiency of those processing mechanisms underlying such effects. For example, in the context of a standard Simon task, varying the local repetition probability of the irrelevant attribute does provide probabilistic information about the stimulus location in the next trial, not unlike in principle to informative central location precues (see, e.g., Jongen & Smulders, 2007; Lu & Proctor, 1995).

Second, recent research on the sequential congruency effect (for recent review, see Duthoo et al., 2014; Hommel, 2011; Pfister et al., 2013; Puccioni & Vallesi, 2012) supports the view that the local repetition probability of higher-order attributes which are unrelated to the response per se can have systematic effects, too. For example, Jimenéz and Méndez (2013) studied the sequential congruency effect in a Stroop task by varying the local repetition probability of a trial to be congruent or incongruent, without altering the overall probability (0.50) of congruent vs. incongruent trials (for a related study using explicit cues indicating the congruency status of the next trial, see Wühr & Kunde, 2008). In their Experiment 2A the congruency status (i.e., congruent vs. incongruent) of the previous trial was

repeated with a local probability of $\pi = 0.70$. Compared to a neutral $\pi = 0.50$ condition (their Experiment 1), this led to an increase of the congruency effect following congruent trials, and to a decrease of the congruency effect following incongruent trials. Jimenéz and Méndez (2013) account for this effect by assuming that participants form expectancies on the basis of learned design contingencies. Thus, following a congruent $n-1$ trial, participants tend to expect another congruent trial which further decreases their congruent trial n RT and increases their incongruent trial n RT, producing an overall increase of the congruency effect in trial n. In contrast, following an incongruent $n-1$ trial, participants tend to expect another incongruent trial which decreases their incongruent trial n RT and increases their congruent trial n RT, thus producing a decrease of the congruency effect in trial n. As expected, complementary results were observed for a $\pi = 0.30$ condition (Experiment 2B) which led to a reduction (an increase) of the congruency effect following a congruent (an incongruent) $n-1$ trial.

The present experiments

The detailed findings just reviewed directly relate to the issues addressed in the present study, and several critical features should be noted in relation to the experiments reported below. First, the RSI of 0 ms, chosen by Jimenéz and Méndez (2013) to minimize the effect of expectancies, is untypical of standard conflict tasks; at present it is an open question whether local repetition probability influences RT at larger RSIs (but see Jimenéz & Méndez, 2014). Furthermore, note that the specific property whose local repetition probability was varied in the study of Jimenéz and Méndez (2013) was not the irrelevant attribute (i.e., the word meaning) as such but rather the congruency status of the trial. To study the effect (if any) of the local repetition probability of response-irrelevant attributes a filtering task design is required in which the irrelevant attribute is locally (from one trial to the next) repeated with a given probability π, independent of the response-relevant attribute, and also independent of the congruency status of the trial. In the absence of extant

studies with these design features, it is an open question whether variations of the local repetition probability of response-irrelevant attributes have any specific differential effects beyond repetition per se at all, and if so whether these effects are similar to those for response-relevant attributes. The present experiments sought to fill in this gap, using two representative and well-investigated paradigms.

In our Experiment 1 we used a standard Simon task, in which the relevant attribute is color, and the irrelevant attribute is location. Our Experiment 2 addressed the same issues in the context of semantic classification, and exploited the well-documented SNARC effect (spatial-numerical association of response codes; Dehaene, Bossini, & Giraux, 1993; Dehaene, 2011; Fias & Fischer, 2005) in which participants indicate the parity (relevant attribute) of single digits by manual responses (for saccadic and pedal responses, see Schwarz & Keus, 2004; Schwarz & Müller, 2006), the irrelevant attribute in this case is whether the digit is numerically small (<5) or large (>5).

For an unambiguous attribution of the effects of varying the local probability of the response-irrelevant attribute it is of central importance to use stimulus sequences which meet several crucial boundary conditions (for a set of systematic studies of contingency learning effects in this context, see Schmidt, 2013; Schmidt & Besner, 2008; Schmidt, Crump, Cheesman, & Besner, 2007; Schmidt & De Houwer, 2012a, b; Schmidt, De Houwer, & Besner, 2010). In particular, we generated stimulus sequences in which overall both levels of the relevant attribute (and thus both responses) were presented equally often, the same was true of the levels of the irrelevant attribute, and also for congruent vs. incongruent trials. Furthermore, in our design repeating vs. alternating the irrelevant attribute was independent of repeating vs. alternating the relevant attribute (and thus the response), and also independent of whether the trial was congruent or not. Finally, in all conditions, the relevant attribute was repeated with a probability of 0.50; the same was true of the congruency status (e.g., Jimenéz & Méndez, 2013).

Experiment 1

Experiment 1 was a visual Simon task, with color as the relevant and location as the irrelevant attribute. The factor of central importance, varied between participants, was the probability π (= 0.25, 0.50, 0.75) that the irrelevant attribute (i.e., location) was repeated from one trial to the next.

Method

Participants. Thirty-six University of Potsdam students, aged 18-33, with normal or corrected-to-normal vision participated in one session of approximately one hour. They either received a payment of €8 or course credit for their participation.

Stimuli and apparatus. The stimuli were circular color (red/green) patches (1.2 deg diameter) presented against a gray background 4 deg to the right or left of a centered fixation cross on a 75 Hz, 1028 × 768 pixel VGA monitor. Participants responded by pressing buttons on custom-built response boxes with the index fingers of their left or right hand; the response boxes were fixed on a table 25 cm to the left and right of the midline.

Procedure. Twelve participants were randomly assigned to each group defined by the probability π (= 0.25, 0.50, 0.75) that the location was repeated from one trial to the next. Following one practice block, in each of ten blocks (comprising three warm-up trials, followed by 72 regular trials) each of the four combinations of color and location was shown 18 times, so that each color and each location occurred 36 times. Color (and thus the response) was repeated/alternated 36 times, the same was true of congruency status. For the π = 0.25 (0.50, 0.75) group, location was repeated in 18 (36, 54) of the 72 trials in each block. Sequences were generated such that location repetitions were independent of both color and color repetitions, and also of congruency status and congruency status repetitions.

The task was to indicate, as quickly as possible, the color of the stimulus by a button press. Six participants in each group worked under each of both color-to-

response mappings (this variable had no main or interaction effect, and is ignored in the following). Each trial started with the presentation of a central fixation cross. After a delay of 500 ms the stimulus was presented until a button was pressed; the next trial then started 1000 ms after the response.

All effects were tested at a significance level of $\alpha < .05$. For a trial to enter into the analysis of mean RT the response had to be correct, and be preceded by a correct response (exclusion rate of 8.4 per cent); also, RTs outside the range of 150 ms to 900 ms were excluded (a rate of 0.9 per cent).

Results

Mean RT. We analyzed mean RT by a 3 × 2 × 2 × 2 mixed-model Anova in which the repetition probability of the irrelevant location was a between-subjects factor (3: 0.25, 0.50, 0.75); three binary within-subject factors were defined by whether or not in the trial considered the irrelevant attribute (location) was repeated (2), the relevant attribute (color, and thus the response) was repeated (2), and whether the trial was congruent or incongruent (2).

Overall, response (i.e., color) repetitions (379 ms) were faster than alternations (402 ms; F(1,33) = 37.90, MSE = 1045, η^2=0.54, p < .001; Figure 1, right panel), and congruent trials (382 ms) led to shorter RTs than incongruent trials (399 ms, Simon effect; F(1,33) = 40.11, MSE = 541, η^2=0.55, p < .001; Figure 2). In contrast, repeating (390 ms) vs. alternating (391 ms) the irrelevant location had no main effect on RT (F(1,33) = .28, MSE = 519, η^2=0.01, p < .598), confirming the previous findings reviewed in the Introduction.

The most relevant finding in the present context is the interaction of the local repetition probability π for the irrelevant location with its actual repetition or alternation in a given trial (F(2,33) = 6.94, MSE = 519, η^2=0.30, p < .003), shown in the left panel of Figure 1. For the π = 0.25 group repeating the location increased mean RT by 8 ms, relative to alternating the location. In contrast, for the π = 0.75 group repeating the location decreased mean RT by 15 ms. For the π = 0.50 group

repeating (375 ms) or alternating (372 ms) the location had no reliable effect ($t(11)$ = 1.09, η^2=0.10, p = .30). The group factor defined by the three levels of π neither modified the response repetition main effect (Figure 1, right panel; F(2,33) = 2.43, MSE = 1045, η^2=0.13, p < .104) nor the Simon main effect (Figure 2; F(2,33) = 0.66, MSE = 541, η^2=0.04, p < .523).
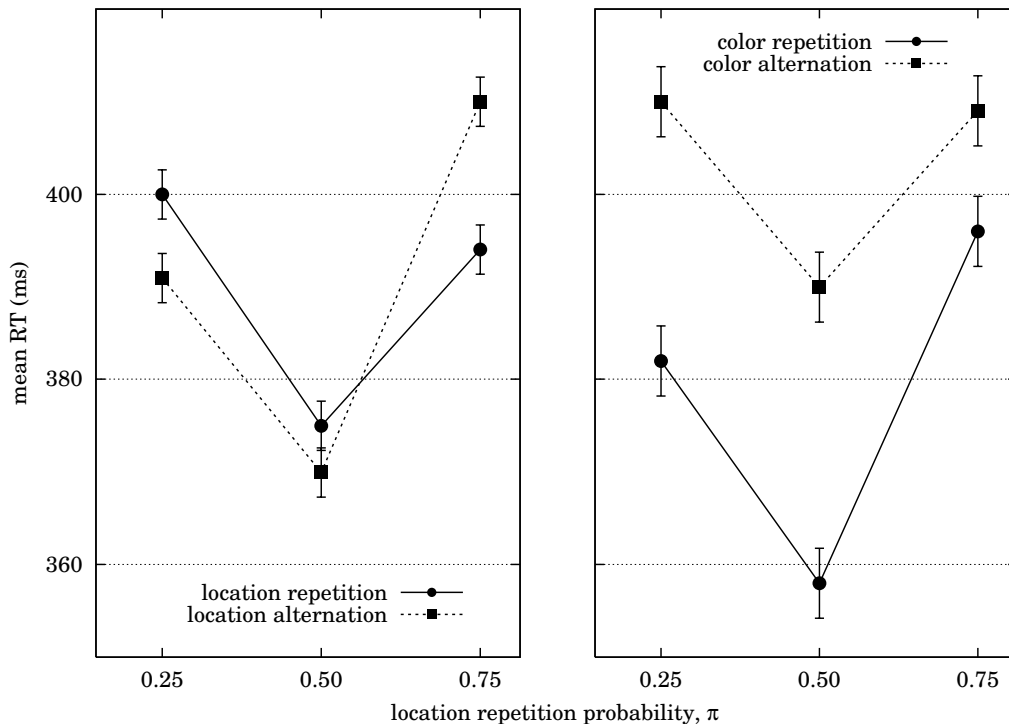

Figure 1. Both panels: mean RT (ordinate in ms) for three groups with local probability of location repetition equal to π = 0.25, 0.50, 0.75 (abscissa). Left: mean RT for location repetition (dots, solid lines) vs. location alternation (squares, dotted lines) trials. Right: mean RT for color (thus, response) repetition (dots, solid lines) vs. color alternation (squares, dotted lines) trials. Error bars show ± 1 s.e.m. (Loftus & Massey, 1994).

Only two further interaction effects were significant, both confirming effects reviewed in the Introduction, and none of which involved the group factor for the levels of π. First, the congruency effect is larger (22 ms) when the location is alternated compared to when it is repeated (13 ms; F(1,33) = 8.92, MSE = 142, η^2=0.21, p < .005). Second, when the location is repeated, color (and thus response) repetitions are 55 ms faster; this turns into an advantage for color (and response) alternations of 8 ms when the location alternates (F(1,33) = 279.73, MSE = 244, η^2=0.89, p

$< .001$). In contrast, we found no significant three-way interaction of the location repetition probability π with the location's actual repetition or alternation in a given trial, and congruency (F(2,33) = 0.11, MSE = 142, η^2=0.00, p $=$ .90).
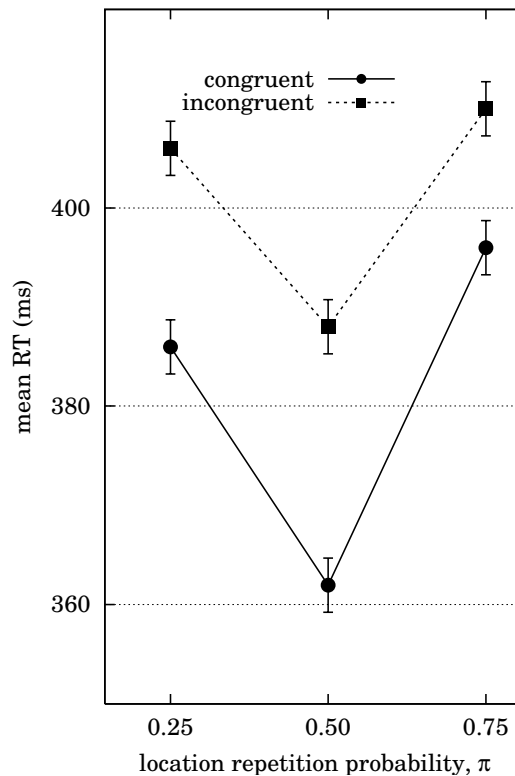

Figure 2. Mean RT (ordinate in ms) for three groups with local probability of location repetition equal to π = 0.25, 0.50, 0.75 (abscissa). Dots and solid lines indicate congruent, squares and dotted lines incongruent trials. Error bars show $\pm$ 1 s.e.m. (Loftus & Massey, 1994).

Error rates. The overall error rate (in per cent) was 4.3. The results of the Anova of error rates mostly conform to those for RT, such that long RTs go with high error rates. Specifically, fewer errors were made in congruent (3.7) than in incongruent (5.0) trials (F(1,33) = 4.78, MSE = 21.50, η^2=0.13, p $<$ .036). Of central importance is the interaction of the irrelevant repetition probability π with repeating vs. alternating the irrelevant location (F(2,33) = 8.98, MSE = 7.26, η^2=0.35, p $<$ .001). This interaction reflects that participants in the π = 0.25 group produced more errors when the irrelevant attribute (location) was repeated (5.4) than when it

alternated (4.2), whereas for the $\pi = 0.75$ group this pattern was exactly reversed (3.7 for repeated vs. 5.8 for alternated locations). As was the case for RT, the $\pi = 0.50$ group showed no reliable error rate effect for repeating (3.3) vs. alternating (3.8) the location.

Only two further interaction effects were significant, none involving the group factor. First, when the color is repated then incongruent trials lead to more errors (5.7) than congruent ones (3.4). However, when the color alternates, the error rates in incongruent (4.3) and congruent trials (4.1) are nearly equal, resulting in an interaction of relevant (i.e., color) repetition with congruency ($F(1,33) = 10.67$, MSE $= 7.52$, $\eta^2=0.24$, p $<$ .003). Second, confirming the analogous effect for mean RTs, when the location is repeated, fewer errors (2.2) are made when color is repeated, too, than when color alternates (6.0). Conversely, when the location alternates, more errors are made for color repetitions (6.8) than for color alternations (2.4), producing a highly reliable interaction ($F(1,33) = 42.91$, MSE $= 28.55$, $\eta^2=0.57$, p $<$ .001).

Discussion

The results of Experiment 1 provide a first demonstration that variations of the local transition probability π of repeating response-irrelevant attributes (location) can have reliable and consistent effects on performance in a standard filtering task. That is, the present results indicate that location is not irrelevant in the sense that expectations about the location of the forthcoming stimulus based on the first-order transition probabilities have no systematic effects on performance (for a discussion of the conditions under which these expectations become explicit, see, e.g., Blais, Harris, Guerrero, & Bunge, 2012; Bugg, Diede, Cohen-Shikora, & Selmeczy, 2015; Schmidt et al., 2007). Specifically, participants in the group with a high local probability of repeating the irrelevant location responded faster and more accurately in trials in which the location is in fact repeated, and a converse pattern was observed in the group for which π was low.

Given that this effect in turn neither modified the response repetition benefit nor the congruency effect, one plausible interpretation, then, is that the local repetition probability for location speeds up an early orienting stage, much like informative central location cues do (e.g., Girardi, Antonucci, & Nico, 2013, 2015). More specifically, the participants in groups π = 0.25 and 0.75 might interpret the location of the current stimulus as a (mostly) valid cue for the stimulus location in the next trial, and benefit from a corresponding attentional orientation whenever the stimulus in fact occurs at the more likely location. Reviewing precueing effects in the Simon task, Lu and Proctor (1995) concluded that "precues influence the magnitude of the Simon effect only when an expectancy for a particular *response* is created ... but not by an expectancy for the particular *location* in which the target stimulus will occur" (p. 187, italics added). Similarly, Wühr & Kunde (2008) observed that in a Simon task participants cannot make use of reliable cues about the congruency of a forthcoming stimulus to change the attentional weights of processing channels for different stimulus dimensions, which is in line with Jimenéz and Méndez (2013, p. 281) who conclude that although "the manipulated contingencies had been effective in changing participants' expectancies concerning the congruency of the next trial, the effects of congruency remained relatively immune to such contingencies". Thus, our own findings and the interpretation given fit in well with related standard findings from the Simon task that valid central location precuing leads to an overall speed benefit, but does not modify the size of the Simon effect.

To assess the generality of our findings, it would be important to extend the basic design used in Experiment 1 in several ways. First, given the findings summarized in the Introduction regarding partial repetition benefits in semantic classification, it seems natural to ask if the effects of the local repetition probability for irrelevant attributes extend beyond simple sensorimotor (e.g., Simon) tasks. More specifically, while the effects observed in Experiment 1 could possibly reflect shifts of spatial attention to the location predicted by the preceding trial, analogous

effects with non-spatial predictability would presumably require a more general explanation.

Second, note that the standard Simon task used in Experiment 1 involves only four different stimuli which means that repeating both the relevant and the irrelevant attribute implies an exact replica of the previous trial, which potentially provides a repetition benefit specific to identical repetitions (see, e.g., Hommel, 2011; Pashler & Baylis, 1991; Schwarz & Ischebeck, 2000).

It is, therefore, conceivable that our findings regarding the critical interaction of the local repetition probability π for the irrelevant location with its actual repetition or alternation in a given trial (left panel of Figure 1) reflects an effect that is closely tied to identical repetitions[1]. To test this conjecture, we run an additional analysis in which all trials were removed in which the stimulus presented was identical (i.e., in location and color) to the stimulus in the preceding trial. In the design of Experiment 1 repeating the relevant attribute color and repeating the irrelevant attribute location were always independent events: in all three π = 0.25/0.50/0.75 groups, there was an equal proportion (i.e., 0.50) of relevant repetitions and relevant alternations both within irrelevant repetitions and within irrelevant alternations. Therefore, the exclusion of identical repetitions maintained the same trial composition between the three $\pi-$groups. Note that excluding identical repetitions does not affect at all the RT means for alternations of the irrelevant attribute, which by definition cannot be identical repetitions. Rather, it only excludes half of all those trials in which the irrelevant attribute is repeated, namely, that half in which the relevant attribute (and thus the response) is also repeated, leaving for irrelevant repetitions only the other half in which the relevant attribute (and thus the response) alternates. As described above, there is a robust interaction of repeating the relevant and the irrelevant attribute such that repeating both, or alternating both, leads to clearly faster RTs than in the two mixed cases. As a result, whereas mean RTs for

---

[1]We thank an anonymous reviewer who suggested this additional analysis.

alternations of the irrelevant attribute remain by definition unchanged by excluding identical repetitions, mean RTs for repetitions of the irrelevant attribute are clearly longer (about 30 ms) after the exclusion of identical repetitions. The crucial finding, however, is that this effect is essentially the same across all three $\pi-$groups, thus producing very nearly the same interaction contrast (F(2,33) = 11.51, MSE = 165, η^2=0.41, p < .0002) of the local repetition probability π for the irrelevant location with its actual repetition or alternation in a given trial as in the original analysis including identical repetitions. Specifically, after excluding identical repetitions the RT advantage of alternations over repetitions of the irrelevant location was maximal (38 ms) for π = 0.25 (where such alternations are expected), decreased to 33 ms for π = 0.50, and was minimal (5 ms) for π = 0.75. Clearly, this analysis suggests that the interaction of π with the actual location repetition or alternation in a given trial found in Experiment 1 cannot be attributed to a processing benefit specific to identical repetitions. Clearly, this conclusion could be further strengthened by using a richer design involving a categorically defined irrelevant attribute such that this attribute can be repeated with non-identical stimuli, even when the relevant attribute (i.e., the response) is repeated, too.

Experiment 2

In Experiment 2 participants classified digits (1-9, excluding 5) according to their parity (relevant dimension), indicating their decision by manual responses (index finger of the left or right hand). Under standard conditions the difference of right-hand RT minus left-hand RT decreases with the numerical magnitude of the digit (the so-called SNARC effect; Dehaene et al., 1993; for review, see Dehaene, 2011; Fias & Fischer, 2005; Schwarz & Keus, 2004). That is, ignoring potential main effects of magnitude and hand, participants respond (relatively) faster to small digits (<5) with their left hand, and to large digits (>5) with their right hand, even

though numerical magnitude per se is a response-irrelevant attribute. A common interpretation is that the size of the SNARC effect reflects the correspondence (or lack thereof) of the internal representation of numerical magnitude in the format of a left-to-right mental number line (Restle, 1970) with the spatial response layout. The aim of Experiment 2 was to evaluate if the findings of Experiment 1 extend to a non-spatial semantic (i.e., parity) classification task, using a design in which even for repetitions of the relevant attribute (i.e., parity) the irrelevant attribute (i.e., digit smaller vs. larger than 5) could be repeated without identical stimulus repetitions (e.g., a 7 followed by a 9).

Method

Participants. Thirty-six University of Potsdam students, aged 18-31, with normal or corrected-to-normal vision participated in one session of approximately one hour. They either received a payment of €8 or course credit for their participation.

Stimuli and apparatus. The stimuli were the digits (Verdana font, height 1.1 deg) 1–9 (excluding 5) presented at fixation; all other technical aspects were identical to Experiment 1.

Procedure. Twelve participants were randomly assigned to each group defined by the probability π (= 0.25, 0.50, 0.75) that the irrelevant numerical magnitude was repeated from one trial to the next. Following one practice block, in each of twelve blocks (comprising three warm-up trials, followed by 64 regular trials) each of the eight digits was shown eight times. Each of the four parity $\times$ magnitude combinations occurred 16 times, so that odd and even digits, and small and large digits occurred 32 times. In congruent trials either a small (<5) digit required a left-hand response or a large (>5) digit required a right-hand response; the other digit-response combinations are incongruent. Parity (and thus the response) was repeated/alternated 32 times, the same was true of congruency status. For the π = 0.25 (0.50, 0.75) group, numerical magnitude (<5 vs. >5) was repeated in

16 (32, 48) of the 64 trials in each block. Sequences were generated such that magnitude repetitions were independent of both parity and parity repetitions, and also of congruency status and congruency status repetitions. There were no trials in which the same digit as in the previous trial was presented.

The task was to indicate, as quickly as possible, the parity of the digit by a button press. Six participants in each group worked under each of both parity-to-response mappings (this variable had no main or interaction effect, and is ignored in the following). Each trial started with the presentation of a central fixation cross which after a delay of 500 ms was replaced with a response-terminated digit; the next trial then started after 1000 ms.

All effects were tested at a significance level of $\alpha < .05$. For a trial to enter into the analysis of mean RT the response had to be correct, and be preceded by a correct response (exclusion rate of 8.8 per cent); also, RTs outside the range of 150 ms to 900 ms were excluded (a rate of 1.7 per cent).

Results

Mean RT. We analyzed mean RT by a 3 × 2 × 2 × 2 mixed-model Anova in which the repetition probability of the irrelevant numerical magnitude was a between-subjects factor (3: 0.25, 0.50, 0.75); three binary within-subject factors were defined by whether or not in the trial considered the irrelevant attribute (magnitude: digit <5 vs. digit >5) was repeated (2), the relevant attribute (parity) was repeated (2), and whether the trial was congruent or incongruent (2).

As shown in Figure 3, response (i.e., parity) repetitions (474 ms) were faster than alternations (484 ms; F(1,33) = 6.17, MSE = 6845, η^2=0.16, p < .018; right panel), and congruent (472 ms) trials led to shorter RTs than incongruent (486 ms) trials (SNARC effect; F(1,33) = 13.90, MSE = 905, η^2=0.30, p < .001; Figure 4). In contrast to Experiment 1, repeating (477 ms) vs. alternating (482 ms) the irrelevant attribute (magnitude) also had a small but reliable main effect, F(1,33) = 20.24, MSE = 88, η^2=0.38, p < .001 (Figure 3, left panel).
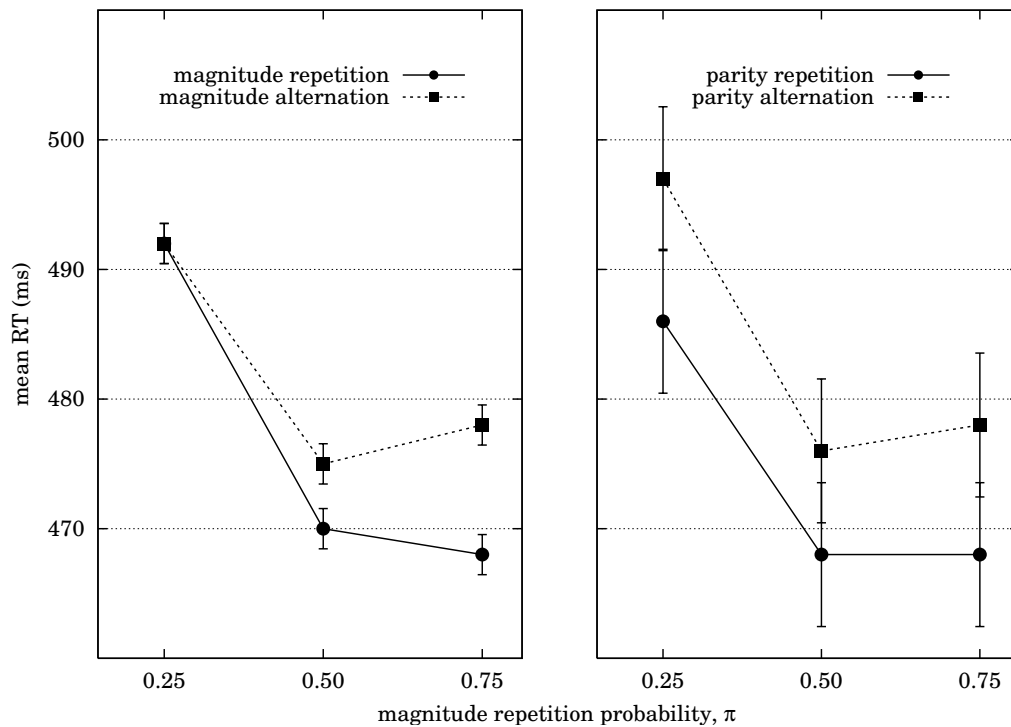
Figure 3. Both panels: mean RT (ordinate in ms) for three groups with local probability of magnitude repetition equal to π = 0.25, 0.50, 0.75 (abscissa). Left: mean RT for magnitude repetition (dots, solid lines) vs. magnitude alternation (squares, dotted lines) trials. Right: mean RT for parity (thus, response) repetition (dots, solid lines) vs. parity alternation (squares, dotted lines) trials. Error bars show $\pm$ 1 s.e.m. (Loftus & Massey, 1994).

Critically, the interaction of the local magnitude repetition probability π with actually repeating vs. alternating magnitude was significant (F(2,33) = 7.34, MSE = 88, η^2=0.31, p < .002). Due to the main effect of magnitude repetition, the exact nature of this interaction takes a slightly different form than in Experiment 1, as shown in the left panel of Figure 3. For the control group (π = 0.50), repeating numerical magnitude decreased mean RT by 5 ms, relative to alternating it, a small but systematic magnitude repetition baseline effect (t(11)=2.26, η^2=0.32, p < 0.045). For the π = 0.75 group, this effect of repeating irrelevant magnitude is further increased by 5 ms to 10 ms (t(11)=4.76, η^2=0.70, p < 0.01), and for π = 0.25 it is decreased by the same amount, so that for this group repetitions and alternations of magnitude produced exactly the same mean RT (t(11)=0.05, η^2=0.00, p < 0.96). The group factor defined by the three levels of π neither modified the response repeti-

tion effect (Figure 3, right panel; F(2,33) = 0.04, MSE = 1110, η^2=0.00, p < .958) nor the SNARC effect Figure 4; F(2,33) = 1.83, MSE = 905, η^2=0.10, p < .176). The only other significant interaction was that when magnitude is repeated, parity (and thus response) repetitions are 15 ms faster; this response repetition benefit reduces to 5 ms when the numerical magnitude alternates (F(1,33) = 16.28, MSE = 130, η^2=0.33, p < .001).
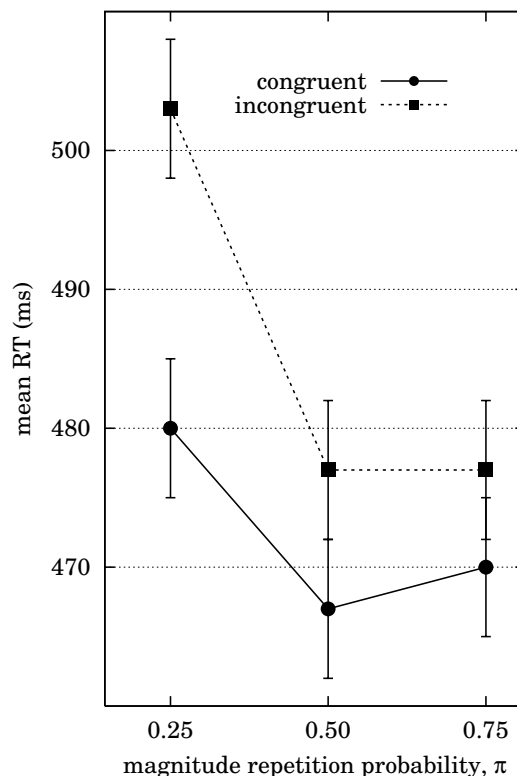

Figure 4. Mean RT (ordinate in ms) for three groups with local probability of location repetition equal to π = 0.25, 0.50, 0.75 (abscissa). Dots and solid lines indicate congruent, squares and dotted lines incongruent trials. Error bars show $\pm$ 1 s.e.m. (Loftus & Massey, 1994).

Error rates. The overall error rate (in per cent) was 4.6. Fewer errors were made when magnitude was repeated (4.3) vs. alternated (5.0; F(1,33) = 4.90, MSE = 1.37, η^2=0.13, p < .034), when parity was repeated (4.0) vs. alternated (5.2; F(1,33) = 5.21, MSE = 18.59, η^2=0.14, p < .029), and in congruent (3.5) vs.

incongruent (5.7) trials (F(1,33) = 11.31, MSE = 30.84, η^2=0.48, p < .002). The only significant interaction shows that fewer errors were made when magnitude and parity were both repeated or both alternated (4.2), compared to repeating one and alternating the other attribute (5.1; F(1,33) = 9.91, MSE = 5.13, η^2=0.23, p < .003). Given that none of the main or interaction effects involving the irrelevant repetition probability π was significant we conclude that the RT findings regarding π cannot be attributed to speed-accuracy trade-offs.

Discussion

The results of Experiment 2 confirm for a semantic classification task that variations of the local probability π of repeating a response-irrelevant attribute (numerical magnitude) can have systematic and reliable effects on performance in a standard non-spatial conflict task. A notable difference to the Simon task (Experiment 1) is that in Experiment 2 repeating the irrelevant attribute per se reduced both RT and error rate. The combined pattern of both Experiments regarding this main effect thus closely parallels the mixed previous findings, reviewed in the Introduction, with no irrelevant location repetition main effects in visual Simon tasks (cf., Notebaert at al., 2001) but irrelevant flanker repetition main effects in letter classification tasks (cf., Kornblum et al., 2002).

The main effect of magnitude repetition indicates that across all levels of π repeating similar numerical magnitudes from one trial to the next provides a genuine benefit which suggests one potential interpretation of the funnel interaction (rather than cross-over, as in Experiment 1, left panel of Figure 1) of magnitude repetition probability π with actually repeating vs. alternating magnitude. According to this interpretation, this interaction represents the additive combination of two separate components. One component reflects a genuine magnitude repetition effect that is independent of expectancies; it therefore arises even for the $\pi = 0.50$ control group for which the stimulus sequence is completely unpredictable, admitting no expectancies above chance level. In contrast, the second component reflects ex-

pectancies based on learned trial contingencies, as were observed in Experiment 1. Specifically, participants in the $\pi = 0.75$ group tend to expect magnitude repetitions: this leads to a benefit, on top of the genuine component, if magnitude is in fact repeated but to a cost if magnitude alternates. Participants in the $\pi = 0.25$ group tend to expect magnitude alternations, with an ensuing benefit if in fact magnitude alternates; however, if contrary to those expectancies magnitude is repeated the cost of the expectancy mismatch is reckoned up against the benefit from the genuine component.

General Discussion

An extensive body of research has established that repeating irrelevant but potentially conflicting stimulus attributes across trials can have reliable effects, under a wide range of well-understood boundary conditions (cf., Kornblum & Stevens, 2002; Puccioni and Vallesi 2012; Zeischka et al., 2010). To this research on sequential effects in filtering paradigms the present study adds as a new form of manipulation the local trial-to-trial transition probability π of repeating a stimulus attribute that is logically and statistically irrelevant (i.e., unpredictive of the response required) yet potentially conflicting.

Local repetition probability (Bertelson, 1961; for reviews, see Jones et al., 2015; Soetens, 1998) has been extensively investigated before in connection with response-relevant attributes, but not in the context of response-irrelevant attributes in filtering paradigms. Because irrelevant local repetition probability was manipulated, events in the current trial provide a probabilistic basis to form expectancies about, for example, the location (Experiment 1) or the magnitude (Experiment 2) of the forthcoming stimulus. Our results show that these expectancies derived from learned design contingencies can influence how fast (Experiments 1 and 2) and accurate (Experiment 1) responses in the forthcoming trial are. Specifically, rela-

tive to the $\pi = 0.50$ control groups working with unpredictive stimulus sequences, in both experiments participants in the $\pi = 0.75$ groups responded faster when the irrelevant attribute was in fact repeated, and a complementary pattern was observed for groups with $\pi = 0.25$. On the other hand, neither the reliable RT benefit associated with repeating the relevant attribute (and thus the response) nor the RT congruency main effects observed in both experiments were differentially modified by variations of π. Note that our results cannot be accounted for by some form of statistical association between the relevant and the irrelevant attribute, nor by an association of events in the present trial with the response or the congruency-status of the previous trial.

For the Simon task (Experiment 1), a natural interpretation is that the irrelevant attribute of the current trial acts much as a (mostly) valid centrally-presented spatial cue for the next trial. Probabilistic knowledge about the location at which a critical stimulus will appear has well-documented facilitatory effects in valid trials, and detrimental effects in invalid trials (e.g., Foley & Schwarz, 1998; Girardi, Antonucci, & Nico, 2013, 2015; Gould, Wolfgang, & Smith, 2007). This view is consistent with previous findings that location precues in Simon tasks produce the usual cost/benefit pattern but do not differentially influence the magnitude of the Simon effect (e.g., Lu & Proctor, 1995; Wühr & Kunde, 2008). Our interpretation also fits in, for example, with the attention-shift account of spatial correspondence effects (for recent review, see Duthoo et al., 2014; Puccioni and Vallesi 2012; Zeischka et al., 2010) according to which shifts of spatial attention contribute to the Simon effect as a key factor.

On this interpretation, why should irrelevant numerical magnitude in Experiment 2 also lead to corresponding effects of irrelevant repetition probability? According to a standard interpretation of number-related congruency effects one format in which we represent numbers is topologically akin to the spatial stimulus layout in a Simon task, the fundamental difference being, of course, that location in

the parity task corresponds to the internal location of a number along an ordered quasi-spatial representation (the "mental number line"; Restle, 1970). In this interpretation, expectancy effects regarding sub-intervals of the mental number line may facilitate the identification of the to-be-presented number (e.g., Schwarz & Ischebeck, 2000; Tan & Dixon, 2011), and thus of its parity, not unlike a valid spatial location cue in a Simon task. Considerable support for this view is, for example, provided by the work of Nieder (e.g., 2005, 2011) who explored basic neuronal mechanisms of numerical magnitude coding. Recording from cells in the intraparietal sulcus and the prefrontal cortex of rhesus monkeys, he identified a numerical distance-dependent gradient of neuronal activation, yielding well-defined neuronal numerosity-tuning functions for single cells. If this distance-dependent mechanism is transiently facilitated by residual activation from the previous trial, then repeating digits of similar magnitude in subsequent trials might facilitate an initial orientation response along the mental number line.

More generally, expectancy effects as induced by local repetition probability offer a promising approach to further explore how stimulus attributes are processed which by themselves are uniformative about the required response but (in typical filtering paradigms) do induce reliable response repetition and/or congruency effects. Two aspects of this approach seem especially relevant in the present context. The first is that changes of the *global* probability of stimulus attributes imply changes of its local first-order transition probabilities, too, whereas the converse is not true (cf., Kornblum, 1968; Miller, 1998). For example, if in a Simon task as used in Experiment 1 in any trial the stimulus is independently presented in the left location with probability 0.80 (0.50), then 100 trials will on average contain 68 (50) first-order location repetitions. This confound makes it difficult to attribute the effects of varying the global probability of irrelevant attributes either to cumulative overall learning effects (due simply to more frequent presentations of one location; cf., Kabata & Matsumoto, 2012; Miller, 1988, 1998) or to more local, expectancy-

related sequential effects. In contrast, the local repetition probability π of irrelevant stimulus attributes can be manipulated without altering the global (overall) probability of that attribute. Second, it is difficult to attribute the local repetition effects of two-valued relevant attributes in two-choice paradigms (such as color in Experiment 1, or parity in Experiment 2) to chronometrically early (stimulus-related) or late (response-related) processing stages, as repetitions of the relevant attribute and of the associated response are perfectly confounded in such designs. In contrast, the local repetition effects of irrelevant attributes obtained in the present study are independent both of the relevant stimulus attribute (and thus the response) and of congruity status, and also of whether or not these attributes have been repeated from the previous trial or not. Conflict paradigms are by definition based on manipulating the relation of relevant and irrelevant stimulus attributes; given that the variation of the local repetition probability of irrelevant attributes avoids the confounds mentioned above, the general design behind the present experiments allows for cleaner conclusions about how expectancy effects modulate the adaptation to inconsistent or contradictory information.

In conclusion, the present experiments demonstrate systematic and reliable effects due to variations of the local trial-to-trial repetition probability of response-irrelevant attributes (location, numerical magnitude) in a spatial (Simon) and in a semantic (SNARC) filtering task. These effects influence performance in ways comparable to central probabilistic location (Experiment 1) or magnitude (Experiment 2) cues: they exhibit the cost-benefit pattern typically associated with such cues, but do not in turn differentially modify other main (response repetition and congruency) effects. More generally, the design used in our experiments provides a novel approach to explore how trial-to-trial expectancies regarding response-irrelevant attributes influence the nature and limits of cognitive control in filtering tasks.

References

Bertelson, P. (1961). Sequential redundancy and speed in a serial two–choice responding task. *Quarterly Journal of Experimental Psychology*, **13**, 90–102.

Blais, C., Harris, M.B.H., Guerrero, J.V., & Bunge, S.A. (2012). Rethinking the role of automaticity in cognitive control. *The Quarterly Journal of Experimental Psychology*, **65**, 268–276.

Botvinick, M.M., Braver, T.S., Barch, D.M. Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Quarterly Journal of Experimental Psychology*, **108**, 624–652.

Bugg, J.M., Diede, N.T., Cohen-Shikora, E.R., & Selmeczy, D. (2015). Expectations and experience: Dissociable bases for cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 1349–1373.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event–related potentials and the additive–factors method. *Journal of Cognitive Neuroscience*, **8**, 47–68.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*, revised and updated edition. New York, NY: Oxford University Press.

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

Duthoo, W., Abrahamse, E.L., Braem, S., Boehler, C.N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: An update. *Frontiers in Psychology*, **5**.

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143–149.

Fias, W., & Fischer, M.H. (2005). Spatial representation of numbers. In: Campbell, J.I.D. (Ed.), *Handbook of mathematical cognition*. New York: Psychology Press. pp. 43–54.

Foley, J., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold vs. contrast (TvC) function for contrast discrimination. *Journal of the Optical Society of America, A*, **15**, 1036–1047.

Frings, C., Rothermund, K., & Wentura, D. (2007). Distractor repetitions retrieve previous responses to targets. *The Quarterly Journal of Experimental Psychology*, **60**, 1367–1377.

Girardi, G., Antonucci, G., & Nico, D. (2013). Cueing spatial attention through timing and probability. *Cortex*, **49**, 211–221.

Girardi, G., Antonucci, G., & Nico, D. (2015). Timing the events of directional Cueing. *Psychological Research*, **79**, 1009–1021.

Gould, I.C., Wolfgang, B.J., & Smith, P.L. (2007). Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. *Journal of Vision*, **7**.

Gratton, G., Coles, M.G.H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, **121**, 480–506.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, **136**, 189–202.

Jentzsch, I., & Leuthold, H. (2005). Response conflict determines sequential effects in serial response time tasks with short response-stimulus intervals. *Journal of Experimental Psychology: Human Perception and Performance*, **31**, 731–748.

Jimenéz, L., & Méndez, A. (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, **39**, 271–284.

Jimenéz, L., & Méndez, A. (2014). Even with time, conflict adaptation is not made of expectancies. *Frontiers in Psychology*, **5**.

Jones, M., Curran, T., Mozer, M.C., & Wilder, M. H. (2015). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, **120**, 628–666.

Jongen, E.M.M., & Smulders, F.T.Y. (2007). Sequence effects in a spatial cueing task: Endogenous orienting is sensitive to orienting in the preceding trial. *Psychological Research*, **71**, 516–523.

Kabata, T., & Matsumoto, E. (2012). Cueing effects of target location probability and repetition. *Vision Research*, **73**, 23–29.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In: R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. New York: Academic Press.

Kleinsorge, T. (1999). Response repetition benefits and costs. *Acta Psychologica*, **103**, 295–310.

Kornblum, S. (1968). Serial-choice reaction time: Inadequacies of the information hypothesis. *Science*, **159**, 432–434.

Kornblum, S., & Stevens, G. (2002). Sequential effects in dimensional overlap: findings and issues. In: W. Prinz & B. Hommel (Eds.), *Common Mechanisms in Perception and Action*, Attention and Performance XIX. London: Oxford University Press.

Leuthold, H., & Sommer, W. (1993). Stimulus presentation rate dissociates sequential effects in event-related potentials and reaction times. *Psychophysiology*, **30**, 510–517.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476–490.

Lu, C., & Proctor, R.W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, **2**, 174–207.

Marcel, T., & Forrin, B. (1974). Naming latency and the repetition of stimulus categories. *Journal of Experimental Psychology*, **103**, 450–460.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163–203.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, **6**, 450–452.

Meuter, R.F.I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* , **40**, 25–40.

Miller, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 453–471.

Miller, J. (1991). The flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: A search for for boundary conditions. *Perception & Psychphysics*, **49**, 270–288.

Miller, J. (1998). Effects of stimulus-response probability on choice reaction time: Evidence from the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 1521–1534.

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, **6**, 177–190.

Nieder, A. (2011). The neural code for number. In S. Dehaene, S. & E.M. Brannon, E.M. (Eds.), *Space, time and number in the brain. Searching for the foundations of mathematical thought.*, ch. 8, (pp. 103–118). London: Academic Press.

Notebaert, W., & Soetens, E. (2003). The influence of irrelevant stimulus changes on stimulus and response repetition effects. *Acta Psychologica*, **112**, 143–156.

Notebaert, W., Soetens, E., & Melis, A. (2001). Sequential analysis of a Simon task – evidence for an attention-shift account. *Psychological Research*, **65**, 170–184.

Pashler, H. (1998). *The psychology of attention*. London: The MIT Press.

Pashler, H., & Baylis, G. (1991). Procedural learning: Intertrial repetition effects in speeded–choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 33–48.

Pfister, R., Schroeder, P.A., & Kunde, W. (2013). SNARC struggles: Instant control over spatial–numerical associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **39**, 1953–1958.

Pratte, M.S., Rouder, J.N., Morey, R.D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception & Psychophysics*, **72**, 2013–2025.

Puccioni, O., & Vallesi, A. (2012). Sequential congruency effects: disentangling priming and conflict adaptation. *Psychological Research*, **76**, 591–600.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, **83**, 274–278.

Rothermund, K., Wentura, D. & De Houwer, J. (2005). Retrieval of Incidental Stimulus-Response Associations as a Source of Negative Priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 482–495.

Schmidt, J.R. (2013). Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, **20**, 615–630.

Schmidt, J.R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 514–523.

Schmidt, J.R., Crump, M.J.C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, **16**, 421–435.

Schmidt, J.R., & De Houwer, J. (2012a). Contingency learning with evaluative stimuli testing the generality of contingency learning in a performance paradigm. *Experimental Psychology*, **59**, 175–182.

Schmidt, J.R., & De Houwer, J. (2012b). Does temporal contiguity moderate contingency learning in a speeded performance task? *The Quarterly Journal of Experimental Psychology*, **65**, 408–425.

Schmidt, J.R., De Houwer, J., & Besner, D. (2010). Contingency learning and un-learning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, **19**, 235–250.

Schwarz, W., & Ischebeck, A. (2000). Sequential effects in number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1606–1621.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number–size interference effects in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 507–522.

Schwarz, W., & Keus, I.M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, **66**, 651–664.

Schwarz, W. & Müller, D. (2006). Spatial associations in number-related tasks: A comparison of manual and pedal responses. *Experimental Psychology*, **53**, 4–15.

Simon, J.R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The "Stroop effect" vs. the "Simon effect". *Acta Psychologica*, **73**, 159–170.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643–662.

Soetens, E. (1998). Localizing sequential effects in serial choice reaction time with the information reduction procedure. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 547–568.

Tan, S., & Dixon, P. (2011). Repetition and the SNARC effect with one- and two-digit numbers. *Canadian Journal of Experimental Psychology*, **65**, 84–97.

Wühr, P., & Ansorge, U. (2005) . Exploring trial-by-trial modulations of the Simon effect. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, **58**, 705–731.

Wühr, P., & Kunde, W. (2008) . Precueing spatial S-R correspondence: Is there regulation of expected response conflict? *Journal of Experimental Psychology: Human Perception and Performanc*, **34**, 872–883.

Zeischka, P., Deroost, N., Maetens, K., & Soetens, E. (2010). Reduced congruency effects only for repeated spatial irrelevant information. *European Journal of Cognitive Psychology*, **22**, 1137–1167.

Chapter 5

General discussion: Noisy numbers

Cos I'd rather be in this noise

celebrate in this noise colder

It means so much to me

Dixon & Busbee (Archive)

The aim of this thesis is to illuminate the processes behind numerical cognition. How do we use the mental representations of numbers, and how are they influenced? The Chapters 2, 3, and 4 described in considerable detail a process model for number comparison, the influence of numerical representations on the perception of physical values and how we use our mental representation of numbers strategically in number-related tasks. Finally, the results will be discussed further due to mental number representations.

Between number representation and response

It is not obvious how any kind of mental number representation in a number comparison task leads to response effects like the numerical distance effect, which was first described by Moyer and Landauer (1967). Surprisingly, it was not found a quantitatively and easily interpretable process model in the literature whose accountability for response performance was shown in complete paired digit comparisons, explaining many aspects of response time, error rate, variance of response time and error response time. This thesis fills this gap and presents a process model with a parsimonious set of interpretable parameters. Further, aspects like error response times, which were usually neglected, were considered.

In number-related studies, error response times were not included in the analyses. Although response time models were proposed, which have in principle the property to account for error response times, they were not tested. For example, Brown and Heathcote (2008) suggested a ballistic model for response times. In this model, error responses were based on a random variable. Nevertheless, they should not been completely unsystematic. In contrast, the model proposed in Chapter 2 predicts error responses in a very systematic manner (e.g., a numerical distance effect for error responses). These predictions were explicitly tested and confirmed. The results suggest that error responses cannot be reduced to pure

coincidence or not task-related explanations as attention. Rather, it appears to be more inherent in the decision process that sometimes an error has to occur while solving the task in an efficient manner.

The proposed random walk model is based on the assumption of noisy but sorted according to numeric value mental representations of magnitudes (e.g., Nieder, 2005). It describes how these representations are processed. In addition, different scales of number representations were tested. In contrast to a simple linear scale, a logarithmic scale was suggested, in which all decisions depend only on the ratio of the two presented digits. One could argue that at least a third model is missed, a linear scale model but variability increasing also linear with increasing magnitude. As illustrated in Figure 1, in the sense of task difficulty at number comparison it is not possible to discriminate this model from a logarithmic scale model (e.g., Dehaene, 2007). Remarkably, both models produce performance predictions for number comparison tasks that only depended on the ratio of the two compared numbers – the perceived distance.

Cohen and Quinlan (2016) reported computational random walk models to account for aspects of number comparison and fit them to the distribution of response times. Surprisingly, although these models used more parameters than the double exponential model reported in Chapter 2 (and explained less data points), the computational random walk models were successfully fitted only with an additional assumption of an uncertainty for the encoding process of each presented number. In addition, they prioritized the simple linear scale model. Possibly, these different results in contrast to the results reported in this thesis can be explained with the quite different task Cohen and Quinlan (2016) used, a same/different task that contained only eight digit pairs. The random walk model proposed in Chapter 2 showed a considerable fit to the observed data with a complete digit pair comparison (72 pairs) without additional assumptions about the encoding process. However, the non-decision component is much larger than the decision component of
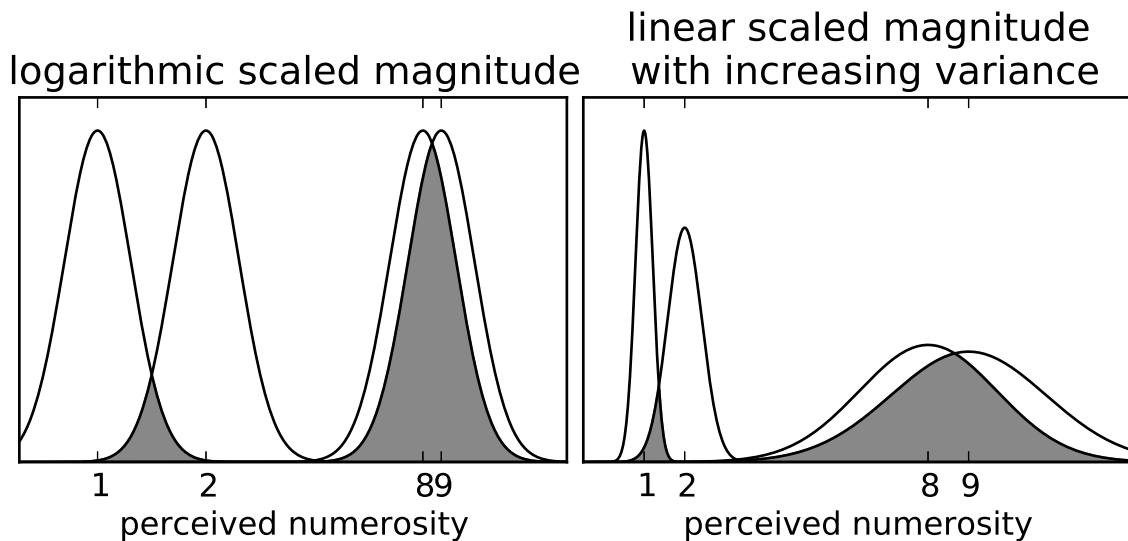
Figure 1. Noisy number representations of perceived magnitude. Both panels show perceptual distributions for 1, 2, 8 and 9. The gray areas illustrate the equality of neighbored numbers. In the left panel, the represented numerical distance between the numbers decreases with increasing magnitude and the variance is unchanged, namely a logarithmic scale. In the right panel, the represented numerical distance is unchanged for neighbored numbers but the variance increases with increasing magnitude, namely a linear scale (with increasing variance). In both panels the overlap of neighbored magnitudes increases with increasing magnitude.

the whole process. It seems worthwhile to examine this component in more detail. It should be possible to integrate two additional parameters to account for manipulations in the encoding process (e.g., adding different visual noise to the stimuli) and in the actual response process (e.g., different effectors: hand vs. foot). Therefore, with this extended model it should be possible to estimate the ratio between the encoding and the actual response part of the non-decision component.

Can the random walk model be more generalized or used for performance predictions in other tasks? Moeller, Huber, Nuerk, and Willmes (2011) used a computational model to investigate two-digit number processing. They prioritize the strictly decomposed model to explain observed data in a two-digit number comparison task. The present random walk model as a quantitative model should be extended to two-digit number processing to validate or refuse this finding. Schwarz

and Eiselt (2012) reported a numerical distance effect in visual search. Godwin, Hout, and Menneer (2014) suggested that these effect depends more on visual similarity than on numerical similarity of the presented digits. Schwarz and Miller (2016) proposed a general quantitative process model (GSDT) to account for usual data in visual search. The present random walk model could be integrated in the GSDT model to investigate in more detail the mental number processing in visual search.

The random walk model is a helpful tool to investigate questions in relation to number comparison tasks in more detail with provable predictions. For example, Sekuler and Mierkiewicz (1977) showed a decreasing numerical distance effect from childhood to adolescence. Norris, McGeown, Guerrini, and Castronovo (2015) reported weaker performance of elderly participants compared to younger participants. Overall, both (children and the elderly) responded slower in number comparison tasks than young adults. In terms of the random walk model, three main hypotheses can be described. First, differences between different groups accounted for real differences in the ability to compare numbers. Second, different groups are more or less accurate. Third, these groups are distinct in other task process components beyond the actual decision process. It can be assumed that children differ in contrast to young adults in the ability to compare numbers. The elderly may not differ in this ability and their slower responses depend on aging processes beyond cognitive numerical skills.

The dependence of physical size and mental number representation

The physical size of numbers influences the performance in numerical judgment tasks, the number-size congruency effect (Besner & Coltheart, 1979). In addition, numerical magnitude influences the performance in physical size judgment tasks

(Henik & Tzelgov, 1982). If physical size and numerical magnitude are congruent, then it seems to be easier to solve these tasks. Until present, it is unclear and debated where this effect originated. Walsh (2003, 2015) proposed a theory of magnitude (ATOM), which can account for the number-size congruency effect and implicit suggests an explanation. Physical size and numerical magnitude partially used the same representation systems and processing stages. Henik, Leibovich, Naparstek, Diesendruck, and Rubinsten (2012) extant this theory with a numerical core system that includes all numerosity processing.

The results reported in Chapter 3 are in line with the view that physical size and numerical magnitude share common representation systems. It can be suggested as an early processing stage associated with the actual perception rather then a late processing stage associated with the response process. In contrast, Sobel, Puri, Faulkenberry, and Dague (2017) investigated the number-size congruency effect in a visual search task and reported that physical size, but not numerical magnitude, guides the visual search. They concluded that physical size and numerical magnitude are integrated in a late processing stage. However, Chapter 3 described sensitivity gain in congruent conditions that suggests the integration of physical size and numerical magnitude cannot be very late. In terms of GSDT (Schwarz & Miller, 2016), visual search items are sorted by a guidance process, which is followed by a serial item inspection. It can be assumed that physical size and numerical magnitude are integrated at the individual item inspection, which does not necessarily have to be a late stage. For future research, to investigate this in more detail, a process model is needed that integrates physical size and numerical magnitude.

Schwarz and Ischebeck (2003) proposed a coalesce model. They integrated physical size and numerical magnitude in a diffusion model that can account for the number-size congruency effect. A similar procedure can be used to integrate physical size in the random walk model described in Chapter 2. This extended

model can be used as a process model to investigate the interaction of physical size and numerical magnitude in considerable detail, also in the original two digit task and in a complete digit pair design.

Physical size is not the only dimension that interacts with numerical magnitude. For example, Hartmann and Mast (2017) reported that loudness influences the perception of numerical magnitude and in the opposite direction numerical magnitude influences the perception of loudness. The experimental design described in Chapter 3 can be easily adapted to loudness. Presumably, the sensitivity gain in congruent condition accounts for the loudness–number association rather than simple numerical mediated response biases.

Strategic use of numerical representations

Participants are sensitive to response repetition probabilities independent of over all stimuli occurrence probabilities (Bertelson, 1961). In conflict tasks, participants are also sensible to repetition probability of congruent stimuli (e.g., Jimenéz & Méndez, 2013). In addition, the sensitivity to detect overall occurrence probability of (in)congruent stimuli operates fast due to an implicit learning process. Resulting expectations do not necessarily need explicit cues (e.g., Bugg, Diede, Cohen-Shikora, & Selmeczy, 2015).

A specific numerical conflict task describes a response speed gain for small (large) numbers for left (right) responses in contrast to right (left) responses, the spatial numerical association response code (SNARC) effect (Dehaene, Bossini, & Giraux, 1993). Participants had to judge the parity of digits and the numerical magnitude was actually irrelevant. The SNARC effect depends on congruency of preceding trials, indicating that number–spatial associations can be accessed quickly and quite flexibly (e.g., Pfister, Schroeder, & Kunde, 2013).

Chapter 4 described an experimental design that is suitable to varying the repetition probability of (irrelevant) stimuli attributes independent of overall stimuli occurrence probability. This design is applied to a Simon task and, more importantly for this thesis, to a SNARC task. In a Simon task, participants are able to use explicit cues successfully (Posner, 1980; for a recent study see: e.g., Girardi, Antonucci, & Nico, 2013). One explanation is a spatial attention shift in the more expected direction of the next stimuli.

The results reported in Chapter 4 suggest that participants are able to use information from preceding trials as a kind of cue for the actual trial (similar to explicit cues). Similar to the Simon task, in the SNARC task participants used the information of the numerical magnitude of the preceding trial. What is the main difference between the Simon task and the SNARC task? In the Simon task, participants perform real spatial shifts of attention whereas in the SNARC task they perform mental (spatial) shifts of attention. Digits can also cause real spatial shifts of attention (Fischer, Castel, Dodd, & Pratt, 2003) but it seems that these numerically initiated shifts of attention are not completely automatic (Fattorini, Pinto, Rotondaro, & Doricchi, 2015). However, it is not necessary to use a mental spatial organization to explain the results reported in Chapter 4. Alternatively, a presented number does not exclusively activate the related mental representation. Instead, close number representations are co-activated (see Nieder, 2005). If participants develop expectations about the numerical magnitude in the next trial using the information of the preceding trial, then they could possibly preactivate the expected range of mental number representations. This procedure does not necessarily require a spatial component. Additionally, according to ATOM (Walsh, 2003, 2015) it should be possible to manipulate this preactivation with physical size, for example.

It can be assumed that the described effects in Chapter 4 are more general and not exclusive to conflict tasks. In a visual search task, it was easy to detect target location probability resulting in related performance gains (e.g., Kabata, &

Matsumoto, 2012). The experimental design described in Chapter 4 can easily adapt to visual search tasks. For example, the task could be to find one landolt ring between closed circles. The ring can be opened to the left or to the right, indicating the left respectively the right answer. In addition, the actual irrelevant position of the landolt ring in the display can be left or right in the display. Now, the repetition probability of this position can be varied independent from overall occurrence probability of left and right opened rings and from overall occurrence probability of left and right positions. It is assumable that then occur similar effects as in conflict tasks reported in Chapter 4. In that case, the strategic use of mental number representation would be reflective of a more general strategic process.

Digits are noisier than expected

If a computer were to solve a number comparison task we would not expect any errors. In addition, most effects described in this thesis (e.g., the numerical distance effect or the magnitude effect) would not occur. The computational number "representation" is very exact and actually not sorted according to numeric value. In contrast, mental number representation seems not to be very exact. The perception of quantities is noisy even though it is sorted according to numeric value. This seems also true if we had to deal with digits. The explanations for most experimental results in this thesis are based on the same basic assumptions. First, that numbers are mentally represented in a noisy manner. Second, that these representations are sorted in a size relation. Alternative assumptions can not explain the results in adequate detail. For example, Cohen (2009) proposed that response times in a same/different task depended on physical rather than on numerical similarity. The task used includes only eight digit pairs. The experimental design reported in Chapter 2 includes 72 digit pairs, and many aspects of number comparison are explained with the random walk model based on noisy represen-

tations of digits. Ratcliff, Thompson, and McKoon (2015) successfully fitted a diffusion model to numerosity and number discrimination tasks. This diffusion model is based on noisy number representation too. Third, interactions with other numerical values, like physical size and strategic uses of numerical representations, are not explained to a satisfying extent without noisy number representations. Therefore, digits activate these representations that can be seen as parts of the meaning of digits.

The variety of presented tasks and reported (numeric) performance in this thesis is in line with the view that we possess a basic ability to understand numbers, known as the number sense (Dehaene, 2011). We use that sense very easily and adapt it to the required task. In addition, we develop a system to communicate number related information.

Symbols represent a great deal of information in a very compact way. Some symbols include more information than formally can be seen implicit in their visual structure. Digits are quite meaningless without related mental representations. It seems that they activate a complex representation system which give us a sense to deal with them. We use special kinds of symbols — digits — to express numerical information in a very compact and compressed manner.

References

Bertelson, P. (1961). Sequential redundancy and speed in a serial two–choice responding task. *Quarterly Journal of Experimental Psychology*, **13**, 90–102.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467–472.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, **57**, 153–178.

Bugg, J.M., Diede, N.T., Cohen-Shikora, E.R., & Selmeczy, D. (2015). Expectations and experience: Dissociable bases for cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 1349–1373.

Cohen, D.J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review*, **16**, 332–336.

Cohen, D.J., & Quinlan, P.T. (2016). How numbers mean: Comparing random walk models of numerical cognition varying both encoding processes and underlying quantity representations. *Cognitive Psychology*, **91**, 63–81.

Dehaene, S.. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Haggard, P., & Rossetti, Y. (Eds.), *Attention & Performance XXII. Sensori-motor foundations of higher cognition* (pp. 527-574). Cambridge, Mass.: Harvard University Press.

Dehaene, S. (2011) *The number sense: How the mind creates mathematics, revised and updated edition*. New York, NY: Oxford University Press.

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

Fattorini, E., Pinto, M., Rotondaro, F., & Doricchi, F. (2015). Perceiving numbers does not cause automatic shifts of spatial attention. *Cortex*, **73**, 298–316.

Fischer, M.H., Castel, A.D., Dodd, M.D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, **6**, 555–556.

Girardi, G., Antonucci, G., & Nico, D. (2013). Cueing spatial attention through timing and probability. *Cortex*, **49**, 211–221.

Godwin, H., Hout, M.C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review*, **21**, 689–695.

Hartmann, M., & Mast, F.W. (2017). Loudness counts: Interactions between loudness, number magnitude, and space. *The Quarterly Journal of Experimental Psychology*, **70**, 1305–1322.

Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., & Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Frontiers in Human Neuroscience*, **5**, 186.

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

Jimenéz, L., & Méndez, A. (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, **39**, 271–284.

Kabata, T., & Matsumoto, E. (2012). Cueing effects of target location probability and repetition. *Vision Research*, **73**, 23–29.

Moeller, K., Huber, S., Nuerk, H.-C., & Willmes, K. (2011). Two-digit number processing: holistic, decomposed or hybrid? A computational modelling approach. *Psychological Research*, **75**, 290–306.

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

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, **6**, 177–190.

Norris, J.E., McGeown, W.J., Guerrini, C., & Castronovo, J. (2015). Aging and the number sense: preserved basic non-symbolic numerical processing and enhanced basic symbolic processing. *Frontiers in Psychology*, **6**.

Pfister, R., Schroeder, P.A., & Kunde, W. (2013). SNARC struggles: Instant control over spatial–numerical associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **39**, 1953–1958.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3–25.

Ratcliff, R., Thompson, C.A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition*, **137**, 115–136.

Schwarz, W., & Eiselt, A.K. (2012). Numerical distance effects in visual search. *Attention, Perception & Psychophysics*, **74**, 1098–1103.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number–size interference effects in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 507–522.

Schwarz, W., & Miller, J. (2016). GSDT: An integrative model of visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **42**, 1654–1675.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, **48**, 630–633.

Sobel, K.V., Puri, A.M., Faulkenberry, T.J., & Dague, T.D. (2017). Visual search for conjunctions of physical and numerical size shows that they are processed independently. *Journal of Experimental Psychology: Human Perception & Performance*, **43**, 444–453.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, **7**, 483–488.

Walsh, V. (2015). A theory of magnitude: The parts that sum to number. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 552565). Oxford, UK: Oxford University Press.