Title
Benford's Law from a Developmental Perspective

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Wünsch, Sophia
Vollmeyer, Regina
Burns, Bruce

Publication Date
2024

Benford's Law from a Developmental Perspective

Sophia Wünsch (sophiawuensch@gmail.com)
Regina Vollmeyer (r.vollmeyer@paed.psych.uni-frankfurt.de)
Goethe University, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany
Bruce D. Burns (bruce.burns@sydney.edu.au)
School of Psychology, The University of Sydney, 2006 NSW, Australia

Abstract

When adults estimate meaningful numbers their distribution of first-digits is strongly biased towards Benford's Law. Insight into why this bias emerges could be gained by examining when it emerges in children. Three hypotheses were formulated: the Representation Hypothesis predicted this distribution can be found in all grades; the Integration Hypothesis predicted a leap in Benford bias from Grade 3 to 4 due to increased mathematical knowledge; and the Distribution Hypothesis proposed a gradual increase across grades due to implicit learning. 151 children in Grades 2 to 4 were asked to estimate numbers based on images and questions. Results showed a strong Benford bias in all three grades but a significant leap from Grade 2 to 3. This was evidence for both the Representation and Integration Hypotheses. Therefore, Benford bias may develop in children due to how they represent numbers, or develop complex mathematical processes, or perhaps some combination of these.

Keywords: Benford's law; estimation; mathematical development; first-digits.

Introduction

People's behavior can follow laws without them even being aware of their existence. Benford's law regarding the distribution of first-digits in data is an example of this because it has been found to fit to adults' estimation of numbers (Burns, 2020; Burns & Krygier, 2015; Chi & Burns, 2022). How and when this develops in children, however, has not been investigated. Thus, the aim of this paper is to make a first step towards understanding Benford's law from a developmental perspective.

Benford's Law

Benford (1938) found that the first-digits of many naturally occurring data approximate a logarithmic distribution. According to this phenomenon, now known as Benford's law, the probability that 1 is the first-digit is 30.1%, that 2 is the first-digit is 17.6%, that 3 is the first-digit is 12.5%, decreasing in probability until 9 is 4.6%. Expressed in a formula, the first-digit d from 1 to 9 occurs with the probability P as follows:

$$P(d) = log10(d + 1) - log10(d)$$

A broad variety of data are a good fit to Benford's law; for example, masses of extrasolar planets and greenhouse gases emissions per country (Joannes-Boyau et al., 2015), taxes (Nigrini, 2012), human electroencephalographic data (Kreuzer et al., 2014) and bots (Mocnik, 2021). It is important to note that in real-life examples the distributions will never exactly fit Benford's law but can only approach them (Smith, 1999).

The question of how Benford's law can be explained seems to have no complete explanation (Berger & Hill, 2011). Hill (1995) offered an approach based on a statistical derivation. He stated: "If distributions are selected at random (in any 'unbiased' way) and random samples are then taken from each of these distributions, the significant [first] digits of the combined sample will converge to the logarithmic (Benford) distribution" (Hill, p. 354). This has similarities to the Central Limit Theorem (CLT), Hill therefore called it the Log-Limit Law for Significant Digits (i.e., first-digits). Smith (1999, p. 721) summarized that the Central Limit Theorem describes how *adding* many random numbers produces a normal distribution whereas multiplying such numbers produces a log-normal distribution.

Not all data will converge to Benford's law, a number of authors (e.g., Fewster, 2009) argue that data should be unbounded, or at least span several orders of magnitude. For example, adults' heights have a narrow range so would not be expected to fit to Benford's law. One implication of this is that data sets consisting of only one or two digit numbers may not be good fits to Benford's law.

Benford's Law as a Behavioral Phenomenon. The ubiquity of Benford's law naturally gives rise to the question: Do humans generate numbers that follow Benford's law? Burns and Krygier (2015) summarized the early attempts to test this question which seemed to yield a negative answer.

The first positive behavioral evidence for Benford's law was Diekmann (2007). He first found that unstandardized regression coefficients reported in journals converged to Benford's law, then asked students to estimate regression coefficients for presented hypotheses. The resulting data approximated Benford's law. However, the sample sizes for Diekmann's two studies were small (n = 10, n = 13).

Burns and Krygier (2015; see also Burns, 2009) asked participants in two studies to answer nine general knowledge questions. The correct numerical answers had been chosen such that each first-digit 1-9 appeared once for correct answers. In both studies the first-digits of participants' answers approximated Benford's law, except that the frequency of first-digit 9 was higher than expected. The authors emphasized that the data did not perfectly fit to

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Benford's law, not surprisingly there was evidence of influences of other heuristics. For example, the elevated digit-5 frequency could be evidence of a bias towards choosing numbers that are halfway between magnitudes. Therefore, Burns (2020) suggested people have a "Benford bias" that distorts estimated numbers towards Benford's law rather than producing perfect fits.

To explain the failure of earlier behavioral studies to find evidence of a Benford bias, Burns and Krygier (2015) pointed out that the earlier studies had asked participants to generate random numbers. Both their questions and Diekmann's (2007) had been meaningful and potentially calculable even if participants had no idea of the correct answers.

Possible Explanations for Benford Bias. Burns and Krygier (2015) proposed that people's approximate fit to Benford's law could be due to them being exposed to such a statistical relationship throughout their lives and thus implicitly learning the pattern from the environment. The large amount of evidence that people can implicitly learn relationships they repeatedly encounter in their environment (e.g., Bargh & Ferguson, 2000) makes this seem a very plausible explanation. So, their second study included a selection task in which nine possible answers to choose from were presented and participants chose one. They hypothesized that if participants had learnt that some first-digits were more common than others then they should favor answers with lower first-digits. The resulting first-digit distribution was essentially flat in contrast to the data from the generation task. Chi and Burns (2022) reinvestigated the comparison between generation and selection task. They tested what they referred to as the Recognition Hypothesis by giving participants a choice between two answers to the general knowledge questions which had difference first-digits. Participants were asked to choose which answer they thought was more likely to be the correct, but the first-digit again had no effect on their responses.

An alternative explanation was that fit to Benford's law is a consequence of the process by which people estimate unknown numbers (Burns & Krygier, 2015; Chi & Burns, 2022). When participants generate a numerical answer to a question they don't know the answer to then it is plausible that they draw upon different pieces of their own knowledge that they think may be relevant, and then combine that information is some way. Such a process may be combining effectively random numbers from different distributions, and if this was the case then it would parallel Hill's (1995) proposal for why for many data sets first-digit distributions approximate Benford's law. As already outlined, Hill suggested that Benford's law may emerge in data that results from combining random selections from a mixture of distributions.

Burns and Krygier's (2015) proposal explained why the generated numbers had to be meaningful, because they imply that in some sense participants were generating Benford bias as a consequence of an attempt at calculation. If this is the case, then mathematical knowledge may play a part in the strength of this bias. Therefore, evidence regarding at what age Benford bias emerges could be important for illuminating why this phenomenon exists. However, first we will review what is known about children's development of estimation of numbers.

Development of Numerosity and Estimating

Some concept of number is present very early in our lives. Already six-month-old infants can discriminate numerosities of 16 versus 8 dots (Xu & Arriaga, 2007). This is an example of what Dehaene (2001) called *number sense* which he considered to be intuitions or basic abilities to "quickly understand, approximate, and manipulate numerical quantities" (p. 16) which resonated as a widely spread concept. Dehaene (1997, 2001) took a biologically determined perspective, referring to evidence like cerebral substrates or basic numerical abilities in animals. The base of his view is that quantities, like in the example of 16 dots, are analogically represented on a mental number line.

To stick with the numerosity example of dots, when estimating these the denominated Approximate Number System (ANS) would be active. The ANS helps to represent approximate quantities, numerosities or magnitudes of sets of objects (Mazzocco, et al., 2011). It is one of the two core systems of numerical representations that are both active from early in infancy onward and observed to be existent in animals as well. In contrast, the second system, the parallel individuation system, serves to represent individual objects in a precise manner (Hyde, 2011; Le Corre & Carey, 2007). The ANS supports the intuitive number sense and improves with age (Mazzocco, et al.). Differentiating between quantities is increasingly difficult when the ratio between the quantities is smaller. This ratio follows the Weber-Fechner Law which is a logarithmic function.

More sophisticated models of the ANS have been developed to explain adult estimation of the number of objects in a visual display (see Brockbank et al., 2022; Cheyette & Piantadosi, 2019). However, these models address estimations of less than 100 visual objects, so how well they can apply to estimations beyond 100 is unclear.

Number Line Estimation Task. When symbolic numbers (e.g., the Arabic digit 9) are learnt, they can be translated into the internal mental representation of their magnitude (Siegler et al., 2009). Once again, rich empirical results found that children at first represent the corresponding magnitude on a logarithmic scale. In the course of development, the mental representation depends on age and numerical range (Siegler & Braithwaite, 2017). In fact, a variety of studies with children and adults (e.g., Berteletti et al., 2010; Siegler et al.) have described a shift from a logarithmic towards a linear representation with age.

The development of the logarithmic-to-linear shift has often been studied with the help of the number-line estimation (NLE) task in which the participants are asked to estimate the location of a presented Arabic numeral on a number line and vice versa (Siegler & Opfer, 2003; Siegler et al., 2009). Relying on a logarithmic representation would mean that small numbers take more space on the number line than large numbers and are thus marked too far on the right (Berteletti et al., 2010). The logarithmic-to-linear shift is dependent on the number range as well. On a range from 0 to 100, children in kindergarten showed the best fit to a logarithmic representation whereas second graders displayed a linear one. The same development happened on a 0 to 1,000 range from second to fourth grade (Siegler et al., 2009). Finally, on a 0 to 100,000 range the logarithmic-to-linear shift became visible between third graders and adults (Thompson & Opfer, 2008). Siegler and Opfer (2003) concluded the existence of multiple representations instead of only one model.

Another notable finding concerning the NLE task is the left digit effect. In the study of Lai et al. (2018), the hundreds digits, like in the case of 398 versus 401, had a strong influence on where to place the number on the line. Consequently, not only the magnitude representation but also the left digit had an impact. It seems to emerge at the age of about 7 and remains in adulthood but with an even stronger effect in children (Lai et al.; Williams et al, 2021). An everyday example of the left digit effect is the different perception of prices like 4.99€ versus 5€ (Lai et al.).

Therefore, we can see that both first-digits and logarithmic relationships may be important developmentally. However, these studies have all examined the representation of numbers rather than their generation. Whether they provide insight into the first-digit logarithmic relationship described by Benford's law for generated numbers remains to be seen.

Formal Mathematics Learning in Elementary School. The number sense is seen as the basis for later formal mathematic knowledge (Dehaene, 2001). This link raises the question of what exactly is taught in formal math education. Examination of the curricula of elementary schools can provide information about the content. Special attention here is put upon the curricula in the Germany federal states of Hesse (Hessisches Kultusministerium [HKM], 2011) and Rhineland-Palatinate (Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur [MBWWK], 2014), where the study in this paper took place.

As stated in the curriculum (MBWWK, 2014), in Grade 1 and 2 different units of measurement are learnt, such as length (e.g., cm), time (e.g., minutes) or money (e.g., €) which is completed in Grades 3 and 4 with area (e.g., cm^2), volume (e.g., l or m^3) and mass (e.g., kg). Moreover, after the second grade the number range goes up to either 100 or 1,000 depending on the federal state and is expanded up to 1,000,000 when finishing fourth grade (HKM, 2011; MBWWK, 2014). Further, arithmetic operations include addition, subtraction, multiplication and division in the first two grades; and is deepened afterwards (HKM, 2011). Importantly, quantity comparisons and estimation are topics in all grades and even the youngest pupils learn to mentally cluster objects in order to facilitate estimation.

In conclusion, the first two grades provide basic knowledge of understanding of numbers, quantities, patterns, data and space on the one hand and process-related competencies like communicating or problem-solving on the other hand (MBWWK, 2014). This is considerably deepened in Grades 3 and 4 into more complex and solid developing competencies. Important for later considerations is that the number range in second grade is 0-100 or 0-1,000 and in third and fourth grade 0-1,000,000.

Research Gap and Hypotheses

We know that by the time children become adults their number estimates show a Benford bias, but it is an open question when this bias emerges in children. To determine how fundamental Benford bias is, the younger the children we could study the better. However, they have to have a good enough understanding of number to answer multi-digit numerical questions. Therefore, it seems to be most meaningful to investigate elementary school children since their basic numerical understanding as well as formal mathematical knowledge is growing fast, which allows them to give solid answers to advanced estimation questions. Therefore, the research question was whether second, third and fourth graders show evidence of a Benford bias, and if so, is there a difference between grades? We tested three possible hypotheses informed by what we have outlined regarding children's development of an understanding of numbers, formal mathematics, statistical patterns and the process of estimation.

Representation Hypothesis. The first hypothesis is that Benford bias is present as soon as children can use numbers. This could be due to the logarithmic representations of numbers that children appear to have, such as those suggested by NLE tasks. This predicts that children in second, third and fourth grade will be able to generate meaningful data in estimation tasks which shows a Benford bias. Therefore, the Representation Hypothesis states that Benford bias will be present in all three grades, and to an equal degree.

Integration Hypothesis. The second hypothesis is linked to acquiring mathematical skills at school which are incorporated into the estimation process. Competencies such as those described above in the curricula are deepened considerably in third and fourth grade when children learn to solve more complex mathematics problems. These include broadening their number range, arithmetic, and integrating multiple sources of information. This allows the children to generate numbers of a quality apt for Benford's law. The mixture of distributions theorem described in the previous chapter might apply here. Given that the study took place at the beginning of the school year, only fourth graders have undergone this process. Therefore, the Integration Hypothesis predicts Benford bias will only be found in fourth grade.

Distribution Hypothesis. The third hypothesis takes another perspective derived from implicit learning. This suggests that patterns in the environment can be learnt unconsciously and automatically. Since the distribution of Benford's law is all around us, it is gradually absorbed. The older the children the more often they are exposed to the distribution of Benford's law in the environment. Consequently, the Distribution Hypothesis foresees that

Benford bias emerges gradually over the grades. Although Burns and Krygier's (2015) and Chi and Burns's (2022) attempts to test what they referred to as the Recognition Hypothesis did not find support, implicit learning is a plausible enough explanation for Benford bias to justify further examination.

Although each of these three hypotheses make distinct predictions based on different processes, it is possible that more than one of the hypothesized processes influences children's number estimations. If that is the case, then more complex patterns of results may be found.

Method

Participants. The sample was 151 children in Grades 2, 3 and 4 (German 2.-4. Klasse, Grundschule) in German public elementary schools that the research team had links with. Number of participants, age and gender depended on the nature of the participating classes and the available of children during severe COVID-19 restrictions in the first half of the 2021/22 school year. 41 of them were pupils in Grade 2 (M_{age} = 7.5, SD_{age} = 0.6; 25 boys and 16 girls), 49 in Grade 3 (Mage = 8.4, SD_{age}= 0.5; 30 boys and 19 girls) and 61 in Grade 4 (M_{age} = 9.6, SD_{age} = 0.6; 34 boys and 27 girls). The study was preregistered.

Materials. In order to investigate the extent of Benford bias in children, image and verbal estimation questions were used. The intention was to let the children generate numbers over a large range. Easy-to-answer material would not be expected to fit to Benford's law because children would know the right answer. The visual material consisted of eight identical images given to all grades and eight grade-specific images. The images contained a large number of a specified type of objects, for example, foods, flowers, animals, people, vehicles, and books. Figure 1 shows the image used for the question "How birds are in the picture?" The higher the grade the higher the number of objects depicted in the grade-specific images. The aim was to adapt the level of difficulty and the known number range for each grade level according to the curricula. For each image children were asked how many of the relevant object were present in the picture?

Figure 1: Example of an image given to Grade 2 children. They were asked how many birds are in the picture?

In order to include a different type of material which is not visual, 10 verbal questions were given to all grades which had answers that should be large numbers, for example, "How many bristles does a toothbrush have?" Their answers were intended to be difficult to estimate. Giving both verbal knowledge and visual stimuli based estimation tasks could address the question of whether Benford bias arises from modality specific processes.

Procedure. Data collection took place in the children's classrooms. The digital images were shown via a projector or a large video screen. We started with a test image to make sure the children understood the task, then for 10 seconds each the 16 images were presented, first the non-grade-specific set then the grade-specific set. After each image, a white screen was shown for 15 seconds to allow children time to write down their estimated number.

After the visual stimuli were presented, the verbal questions were read aloud and the children were asked to again write down their estimated answer. The questions were based on either numerosity estimation or physical units (e.g., kilogram, meter) which were familiar to the children according to the curriculum. The total duration of the experiment was about 20 minutes per class.

Results

How to Measure Benford Bias. The challenging question of how to measure fit to Benford's law has been responded to by some authors using z-tests or chi-square-tests (e.g., Rauch et al., 2011). However, Burns (2020) takes a critical view of these analyses, in particular that fit to Benford's law is assumed when the null hypothesis (no deviation from Benford's law) fails to be rejected. Therefore, whether or not there is a fit may simply be a question of statistical power.

Burns (2020) argues that it should not be expected that any data will perfectly fit Benford's law, thus with enough power the null hypothesis should always be rejected. Therefore, the aim should be to evaluate the degree to which the data approach Benford's law. Burns (2020) proposes this is best done by focusing on the effect size (η^2) of the calculated linear contrast weighted by the pattern of Benford's law. This effect size can be described as how much variance in the data can be explained by the pattern of Benford's law. Thus, a high effect size is interpreted as a strong Benford bias.

Statistical Analyses. We first counted how often each child used each first-digit in their estimates for their 16 picture questions and 10 verbal questions then converting these frequencies to proportions of all 26 response. Using SPSS software and the GLM procedure, a 3 x 9 mixed design analysis of variance with a linear contrast weighted by the first-digit proportions for Benford's law was conducted on children's first-digit proportions. The between-subjects factor was the grade (second, third, fourth), the within-subjects factor was the nine first-digit proportions. This enabled us to test whether the three grade levels differed in terms of degree of Benford bias, and gave a measure of degree of fit because the effect size η^2 of the weighted linear contrast tells us how much variance in the data was explained by the pattern of Benford's law.

Evidence of Benford bias. Figure 2 indicates that across grades the mean proportions of the first-digits approached Benford's law. Similar to previous studies by Krygier and Burns (2015) a digit-5-peak was observed.

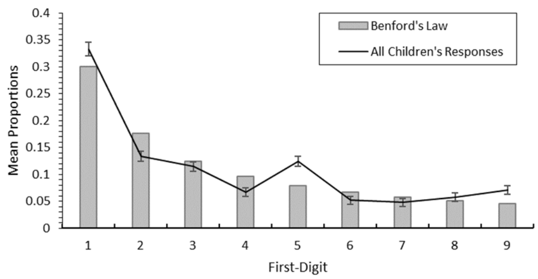

Figure 2: Line Chart of All Children's First-Digit Mean Proportions in Comparison to Benford's Law

Across grades, the linear contrast weighted by Benford's law regarding was statistically significant, $F(1, 148) = 655.20, p < .001, \eta^2 = .816$. Therefore, 81.6% of the variance in the data were explained by the pattern of Benford's law.

Figure 3 differentiates between the three grades and suggests that all grades' first-digit proportions approached Benford's law and had a peak at digit-5. There was a significant interaction between the linear contrast and the grades, $F(2, 148) = 4.21, p = .017, \eta^2 = .054$. Bonferroni-corrected comparisons between the grades identified the only significant difference as between the second and third grade, $F(1, 88) = 6.55, p = .012, \eta^2 = .069$.

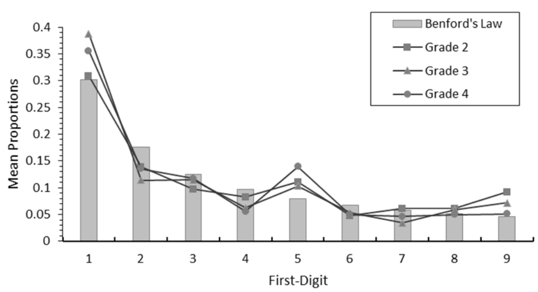

Figure 3: First-Digit Mean Proportions for each Grade in Comparison to Benford's Law.

To determine if Benford bias increased or decreased by grade we estimated the effect size of the weighted linear contrast for each grade. These showed increasing effect size from Grade 2 to 3 as well as 3 to 4: Grade 2, $F(1,40) = 87.53$, $p <.0001, \eta^2 = .686$; Grade 3, $F(1,48) = 237.88, p <.0001, \eta^2 = .832$; Grade 4, $F(1,60) = 422.06, p <.0001, \eta^2 = .876$. Evaluating the three grades separately, the linear contrasts weighted by Benford's law were significant in all three grades with high effect sizes, but there appears to be a leap from Grade 2 to 3.

Further analyses. Exploratory analyses of the other within-subject manipulations found that there were differences due to question type (images or verbal questions), $F(1, 150)= 27.08, p <. 001, \eta^2 = .153$. However, the difference in effect sizes were small, with Bonferroni corrected analyses showing a slightly higher effect size for verbal questions ($\eta^2 = .750$) than for images ($\eta^2 = .719$).

We also analyzed the differences between grade-specific or non-grade-specific images. However, this found no significant difference in Benford bias between image types, $F(1, 150) = 0.16, p = .689, \eta^2 = .001$. That Benford bias was found for both visual stimuli and verbal questions suggests that Benford bias in children is not due to a modality specific process.

Discussion

The main research question was whether elementary school children estimate numbers that approximate Benford's law (as adults do), and, if so, from what grade level? Our results showed that for Grades 2-4 the linear contrast weighted by Benford's law was statistically significant and accounted for a large amount of the variance (81.6%) in the first-digit distribution. The size of this Benford bias was similar to those reported for adults by Burns (2020) and Chi and Burns (2022). Although the effect size remained large across all three grades, there was evidence of an overall increase in effect size across grades particularly from second to third (68.6% to 83.2% explained variance). We can examine how well these results support each of our three hypotheses.

Representation Hypothesis. The observed Benford bias in all three grades supports the Representation Hypothesis which assumed that Benford bias is present from an early age. This could be due to the way children represent numbers as having logarithmic relationships, such as between number and space in the number-line estimation (NLE) task. Siegler and Opfer (2003) also suggest that a logarithmic-to-linear shift increases with age. Weakening this explanation is that it predicts that a Benford bias should have been present in a selection task as well, which was not the case in studies with adults (Burns & Krygier, 2015; Chi & Burns, 2022). However, in this study we did not give a selection task to children, so it is possible that unlike adults, children would show a bias towards smaller first-digits in a selection task. Therefore, it would be of great interest to investigate if children show a Benford bias in selection tasks.

Integration Hypothesis. A leap in Benford bias was predicted by the Integration Hypothesis. This prediction was based on children's ability to handle more complex mathematical problems which includes broadening their number range, arithmetic and integrating multiple sources of information in higher grades. Paralleling the mixture of distributions theorem for Benford's law, that is, combining "random samples from random distributions" (Hill, 1995, p. 358), Burns and Krygier (2015) hypothesized that in an estimation process people combine different information and knowledge. One could assume that using strategies, mathematical operations and multiple sources of knowledge in the estimation process resembles Hill's theorem.

The complex process of estimating requires children to have abilities that are targeted in the curricula (HKM, 2011; MBWWK, 2014) in higher grades, and is most pronounced after completing the third grade. However, the leap in the size of Benford bias was not found from Grades 3 to 4 as predicted but from Grades 2 to 3. This suggests that if the Integration Hypothesis explains Benford bias then the type of information combination suggested involves less calculation than expected.

For estimation of number of visual stimuli it is less obvious what information might be being integrated than it is for the verbal knowledge questions. The images used were not simply dots, so perhaps knowledge about the objects in the image could be integrated. Otherwise perhaps experience or some form of sampling might provide information. Finding Benford bias for both the verbal and visual stimuli suggests that a common process produces the bias.

Distribution Hypothesis. The Distribution Hypothesis predicted a gradual increase in Benford bias across the grades. This was explained by implicit learning from patterns in the environment (e.g., Bargh & Ferguson, 2000), in which Benford's law is ubiquitous. Although an increase in effect size was found, it cannot be described as gradual. In particular, the observed single leap represents an irregularity that does not fit well with the Distribution Hypothesis.

Thus, our results found the most support for the Integration and Representation Hypotheses. One interpretation of this would be that both logarithmic representations and the combining of information is contributing to Benford bias in children in Grades 2-4. However, more research will be needed before we can draw strong conclusion.

Limitations and Future Directions. Data were collected in winter 2021/2022 in the middle of the COVID-19 pandemic. At this time, Germany had experienced a nationwide lock-down since spring accompanied by school closures. Therefore, many participants may not have had the skills that they should according to their grade. This makes it harder to interpret the grade differences, and suggests Grade 2 children had even poorer mathematical skills than we expected.

When collecting data, both images and verbally asked questions were used. The verbal questions produced a slightly higher Benford bias, but the span of the data was wider, which should augment the bias. To substantiate these results, greater attention should be paid to the selection of the material such as its difficulty and the range of numbers. We observed that although children completed the task, they could appear overwhelmed by the difficulty of the questions.

We did not control the first digits of the true answers to our questions. So, to be sure that Benford bias is due to the children giving correct answers we should in the future systematically control the first digit of correct answers.

Finally, the age range should be expanded in future research to depict more comprehensively the developmental pathway of Benford bias.

Conclusion

This study is a first step in investigating Benford bias in elementary school children. The Representation and Integration Hypothesis could partly explain the described results, however, neither explain the observed leap from Grade 2 to 3. Further developmental research could help us better understand why Benford bias exists, and Benford bias could be a tool that helps us understand the development of number estimation and mathematical knowledge in children.

Acknowledgements

We thank Pauline Jung and Sarah Theis for supporting data collection and providing useful discussions.

References

Bargh, J. A., & Ferguson, M. J. (2000). Beyond Behaviorism: On the Automaticity of Higher Mental Processes. *Psychological Bulletin, 126(6*), 925–945.

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society, 78(4),* 551–572.

Berger, A., & Hill, T. P. (2011). Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem. *The Mathematical intelligencer, 33(1),* 85–91.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical Estimation in Pre-schoolers. *Developmental Psychology, 46(2),* 545–551.

Brockbank, E., Barner, D., & Vul, E. (2022). Ongoing dynamic calibration produces unstable number estimates. *Journal of Experimental Psychology: General, 151*, 2092–2114.

Burns, B. D. (2009). Sensitivity to Statistical Regularities: People (Largely) Follow Benford's Law. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31*.

Burns, B. D. (2020). Do People Fit to Benford's Law, or Do They Have a Benford Bias? *42nd Annual Conference of the Cognitive Science Society*.

Burns, B. D., Krygier, J. (2015). Psychology and Benford's Law. In Steven J. Miller (Eds.), *Benford's Law: Theory and Applications,* (pp. 267-275). Princeton: Princeton University Press

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences of the United States of America, 116,* 17729–17734.

Chi, D., Burns, B. D. (2022). Why do People Fit to Benford's Law? - A Test of the Recognition Hypothesis. *Cognitive Science Society, 44*, 3648-3654.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics.* Oxford University Press.

Dehaene, S. (2001). Précis of the Number Sense. *Mind and Language, 16(1),* 16–36.

Diekmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics*, *34(3)*, 321-329.

Fewster, R. M. (2009). A Simple Explanation of Benford's Law. *The American Statistician, 63(1),* 26–32.

Hessisches Kultusministerium (HKM) (2011). *Bildungsstandards und Inhaltsfelder - Das Neue Kerncurriculum für Hessen*: Primarstufe Mathematik.

Hill, T. P. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science, 10(4).*

Hyde, D. C. (2011). Two Systems of Non-Symbolic Numerical Cognition. *Frontiers in Human Neuroscience, 5,* 150.

Joannes-Boyau, R., Bodin, T., Scheffers, A., Sambridge, M., & May, S. M. (2015). Using Benford's Law to Investigate Natural Hazard Dataset Homogeneity. Scientific *Reports,5*.

Kreuzer, M., Jordan, D., Antkowiak, B., Drexler, B., Kochs, E. F., & Schneider, G. (2014). Brain Electrical Activity Obeys Benford's Law. *Anesthesia and Analgesia, 118(1),* 183–191.

Lai, M., Zax, A., & Barth, H. (2018). Digit Identity Influences Numerical Estimation in Children and Adults. *Developmental Science, 21(5),* E12657.

Le Corre, M., & Carey, S. (2007). One, Two, Three, Four, Nothing More: An investigation of the Conceptual Sources of the Verbal Counting Principles. *Cognition, 105(2),* 395–438.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *Plos One, 6(9),* E23749.

Ministerium Für Bildung, Wissenschaft, Weiterbildung Und Kultur (MBWWK) (2014). *Rahmenplan Grundschule*: Teilrahmenplan Mathematik.

Mocnik, F.-B. (2021). Benford's Law and Geographical Information – The Example of OpenStreetMap. *International Journal of Geographical Information Science, 35(9),* 1746–1772.

Nigrini, M. J. (2012). *Benford's Law.* Wiley & Sons, Inc.

Rauch, B., Göttsche, M., Engel, S., & Brähler, G. (2011). Fact and Fiction in EU-Governmental Economic Data. *German Economic Review, 12(3),* 243–255.

Siegler, R. S., & Braithwaite, D. W. (2017). Numerical Development. *Annual Review of Psychology, 68*, 187–213.

Siegler, R. S., & Opfer, J. E. (2003). The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science, 14(3),* 237–243.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The Logarithmic-to-Linear Shift: One Learning Sequence, Many Tasks, Many Time Scales. *Mind, Brain, and Education, 3(3)*, 143–150.

Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing* (2nd Edition). California Technical Pub.

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology, 101*(1), 20–51.

Williams, K., Zax, A., Patalano, A., & Barth, H. (2021). *Left Digit Effects in Numerical Estimation Across Development.* https://doi.org/10.31234/osf.Io/43dq8

Xu, F., & Arriaga, R. I. (2007). Number Discrimination in 10-Month-Old Infants. *British Journal of Developmental Psychology, 25(1),* 103–108.