

Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information

Asmus Leth Olsen*

Abstract

Numerical performance information is increasingly important to political decision-making in the public sector. Some have suggested that biases in citizens' processing of numerical information can be exploited by politicians to skew citizens' perception of performance. I report on an experiment on how citizens evaluate numerical performance information from a public school context. The experiment is conducted with a large and diverse sample of the Danish population (N=1156). The analysis shows a strong leftmost-digit-bias in citizens' evaluation of school grading information. Thus, very small changes in reported average grades, which happen to shift the leftmost grade digit, can lead to very large shifts in citizens' evaluation of performance. The rightmost digit on the grade is almost fully ignored.

Keywords: leftmost-digit-bias, cognitive biases, performance information.

1 Introduction

A democratic society is preserved when the public has reliable ways of knowing whether policies are having the announced or promised effect. Is inflation being brought under control? Is a war of attrition being won? Are defense expenditures buying national security? Numbers, a part of this publicly available political intelligence, consequently contribute to the accountability required of a democracy.

— Kenneth Prewitt (1987, p. 267).

Numerical performance information plays an increasingly important role in political decision-making. Examples include economic statistics, school and hospital rankings, and all other sorts of numerical data aiming to inform citizens about public policy and public sector performance (Dixon, Hood, & Jones, 2008). The disclosure of such data rests on the widely held idea that providing citizens with simple numerical information will improve their ability to make accurate judgments and decisions. This can be in the form of selecting a candidate worthy of ones vote at elections, choosing the right school for the

kids, or finding the best hospital. However, human judgment about numerical values is affected by a number of biases (Miller, 1956; Rosch, 1975; Deahene & Mahler, 1992; Peters, Slovic, Västfjäll, & Mertz, 2008; Smith & Windschitl, 2011; Pope & Simonsohn, 2011). The leftmost-digit-bias constitutes an example: individuals pay too much attention to leftmost digits while partially ignoring digits placed more to the right (Hinrichs, Berie, & Mosell 1982; Poltrock & Schwartz, 1984; Schindler & Kibarian, 1996). While leftmost-digits are more important, the bias implies that digits further to the right are ignored to a greater extent than their numerical value implies.

The leftmost-digit-bias is often attributed to inattention, as individuals use leftmost digits as a judgmental shortcut for processing multi-digit information. Individuals will truncate or use a drop-off mechanism instead of using a more demanding rounding principle when processing numbers from left-to-right (Schindler & Kirby, 1997; Bizer & Schindler, 2005; Thomas & Morwitz, 2005; Korvost & Damian 2008). Evidence from economic markets points to how this inattention to rightmost-digits biases citizens' behavior in response to simple price and quality metrics. In a recent study Lacetera, Pope, & Sydnor (2012) found evidence of the leftmost-digits-bias in car market data. Around changes in odometer values which shift the leftmost digit, they found strong discontinuous changes in prices for sold cars. For instance, a car with an odometer value between 79,900 and 79,999 sells for around \$210 more than cars where the odometer stands between 80,000 and 80,100. For other similar change in odometer value, which leaves the leftmost digit the same, the car price shifts with less than \$10. Su-

The experiment conducted in the article has received funding from the Center for Voting and Parties (CVAP), Department of Political Science, University of Copenhagen. I would like to thank editor Jonathan Baron, Maya Bar-Hillel, and an anonymous referee for very valuable input. The work has also benefited from suggestions provided by Kasper M. Hansen. Any remaining omissions are my own.

Copyright: © 2013. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Department of Political Science, University of Copenhagen, Denmark. Email: ajlo@ifs.ku.dk.

permarkets have long exploited the leftmost-digit-bias by using “odd pricing” such as 9-ending prices (e.g. \$9.99) to reduce the perceived price for consumers (Schindler & Kibarian, 1996; Anderson & Simester, 2003; Bizer & Schindler, 2005). Finally, Pope & Simonsohn (2011) have shown how leftmost-digits can serve as goals. For instance, individuals will invest an extra effort across a number of fields in order to avoid falling below a salient round number.

Here, I ask whether citizens’ biased judgment of multi-digit information is relevant for their evaluation of public sector performance information. Do leftmost digits skew how citizens evaluate numerical performance measures? This question is not only relevant in order to understand if performance information actually leads to better informed citizens. It will also shed some light on the incentives faced by public employees and politicians to manipulate or game numerical performance information (Smith, 1995). Some have argued that numerical performance and policy information can be used for a “politics of digits” (Olsen, 2013). Reelection-minded politicians have strong incentives to increase the perceived numerical value of positive performance information and limit the perception of negative information (Krishna & Slemrod, 2003). That is, politicians can, just like supermarkets, strategically exploit biases in human number-processing (Brunell & Glazer, 2001; Ashworth, Heyndels, & Smolders, 2003). This argument is a special case of the more general idea proposed by McCaffery & Baron (2006, p. 128) that revenue policies should minimize psychic pain while expenditure policies should aim to maximize psychic pleasure (see also McCaffery, 1994). Taxation is here the most straightforward comparison with prices. Politicians’ incentives in the arrangement of taxes are equivalent to those of profit-maximizing retailers: politicians should propose tax digits in order to minimize the political costs of taxation while at the same time maximizing tax revenue (Ashworth, Heyndels, & Smolders, 2003). For instance, Olsen (2013) finds that tax rates follow an odd-pricing logic where 9-ending tax decimals occur up to three times as often as 1-ending taxes.

Drawing on the existing studies in economics and consumer research, the hypothesis in the following is that citizens’ judgment about multi-digit performance information is overly affected by leftmost digits while other digits are partially neglected. I conduct an experiment where (Danish) citizens are provided with hypothetical average grade information from an unnamed school. Numerical educational performance information has been found to affect judgment and choices among both students, parents, teachers, and policy makers (Meredith, 2004; Espeland & Sauder 2007; Bowman & Bastedo, 2009; Pope, 2009). As I will argue later on, the experiment is a rel-

Table 1: Descriptive statistics of the sample

Variable	Pct.
Gender (1=male)	50.5%
Age (mean years)	50.8 (SD=14.9)
Education	
1=High school or less	20.1%
2=Vocational training	24%
3=Short-cycle tertiary	13%
4=Medium-cycle tertiary	27.8%
5=Long-cycle tertiary	15.2%
Kids aged 6–17 (1=yes)	12.3%
School job (1=yes)	16.8%

Note: $N=1156$. CAWI survey in the Danish YouGov panel.

atively hard test of the leftmost-bias: the information is presented without any other attention grabbing information, it consists of a simple two-digit number which has a high degree of familiarity to most citizens. In fact, all Danes have a personal grade average from attending school some point in life. The analysis here points to a strong leftmost-digit-bias in citizens’ evaluation of schools given basic two-digit grade averages. Accordingly, small changes in the assigned grade that happens to shift the leftmost grade digit have large effects on citizens’ subsequent evaluation of school performance. On the other hand, large changes in the average grade within the same leftmost digit have no or small effects on citizens’ evaluation of school performance.

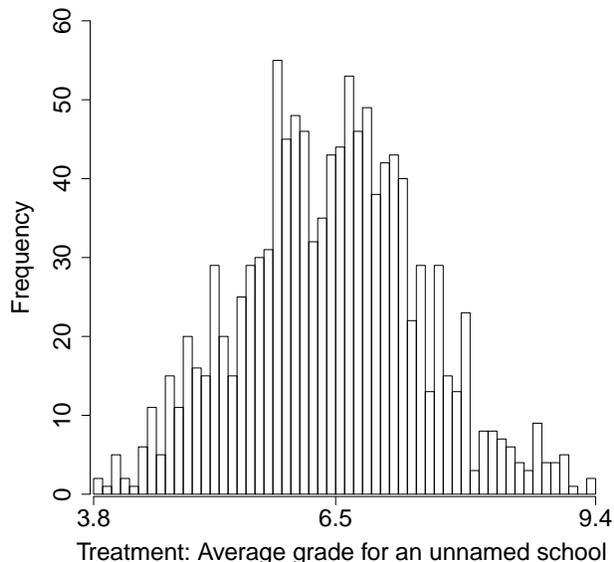
2 Method

2.1 Respondents

Respondents for a survey experiment were recruited via YouGov’s Danish online panel. The sampling frame for the study was restricted to citizens between the age of 18 to 74. The data was collected between the 15th and 22th of October 2012. The response rate was 42%. Table 1 provides an overview of the diversity of the sample used. It shows a highly diverse sample in terms of age, gender, education, and experience with schools (e.g. kids in school age and employment at a school).

In general the survey contained a number of questions dealing with how citizens perceive public services and a number of socio-economic background questions. It also contained a number of experiments. For the experiment presented here 1156 respondents were assigned.

Figure 1: School grade average information drawn from a hypothetical normal distribution ($\mu = 6.5, \sigma = 1.0$).



2.2 Design and procedure

The respondents were asked to evaluate an unnamed school’s performance, given information about its grade average. The respondents were asked the following question on a single screen:

Each year the Ministry of Education releases a grade average for all schools in the country.
 How well do you think this school is doing?
 The school has a grade average of x .

Each respondent was randomly assigned a grade value x drawn from a pre-defined normal distribution. The distribution of grades had an average of 6.5 and a standard deviation of 1.0. The actual distribution for the assigned treatment values is shown in Figure 1.

The distribution of grade averages largely reflects the actual national distribution of grade averages among schools in 2011. In Denmark average grades for each school have been reported publicly for more than 10 years. These grade averages gain a lot of attention from the media and the broader public has a great deal of familiarity with school grade averages. The Danish grade scale is a 7-point scale with values from worst to best being: $-3, 0, 2, 4, 7, 10,$ and 12 .¹ In the experiment, the averages were restricted to one decimal in order to make the treatment a simple two-digit number (e.g., 5.8, 7.4, 8.3, etc.). This corresponds to how the government and the media would normally report grade averages for

¹These grades can be translated into the American scale as: F, Fx, E, D, C, B, and A.

Figure 2: Picture of the response scale: “Very bad” (“Meget dårligt”, 0) to “Very good” (“Meget godt”, 100).



schools. Grade averages for individual students are also rounded to the first decimal on exam transcripts. The grading scale is used at all levels of the Danish educational system from elementary school to the university graduate-level. Furthermore, the high school grade average (with one decimal) will in many instances be the sole determining factor for admission to university. The experiment can therefore be considered a relatively pure test of the leftmost-bias: the information is presented without any other attention grabbing information, it consists of a simple two-digit number and will have a high degree of familiarity to most respondents. It is a natural part of life to think in terms of grade averages with one decimal.

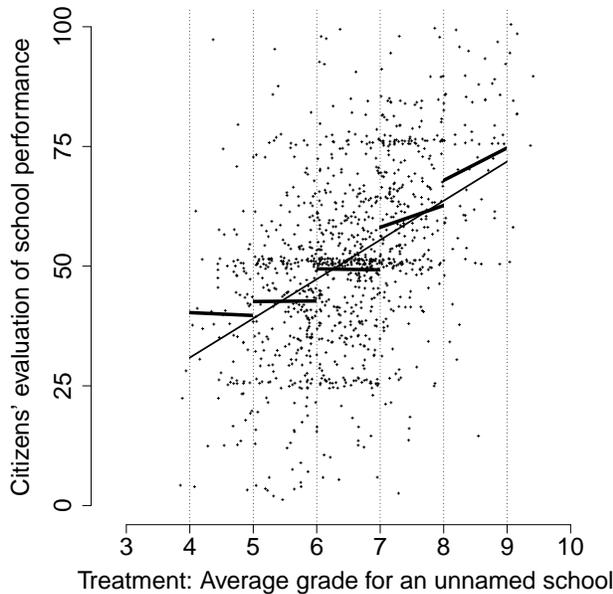
The respondents were instructed to provide their evaluation of the unnamed school on a slider scale ranging from 0 to 100 denoting “very bad” to “very good” performance ($M = 51.36, SD = 18.5$). The underlying score was not visible on the slider, but it was clearly indicated that the respondents could place their answer at any point on the scale. A picture of the response scale is shown in Figure 2. The median response time for the question amounted to 14.34 seconds ($M = 20.95$ seconds, $SD = 70.97$).

3 Results

The first step in the analysis is to test the overall correlation between the treatment grade and citizens’ evaluation of school performance. In Figure 3 we see a strong positive correlation between the average school grade provided and citizens’ evaluation of school performance ($\beta = 8.46, t(1154) = 17.5, p < .001$). However, the correlation is almost solely an effect of changes in the leftmost digit of the school grade and not changes to grade decimals. In the interval 5.0–5.9 the correlation is close to zero and insignificant ($\beta = 0.11, t(263) = 0.03, p = .98$). A similar pattern is found in the range 6.0–6.9 ($\beta = -0.20, t(439) = -0.08, p = .94$) and 7.0–7.9 ($\beta = 4.57, t(282) = 1.27, p = .21$). At face value the correlation seems to become stronger with rising grade averages, but it remains insignificant in the interval 8.0–8.9 ($\beta = 6.79, t(73) = 1.08, p = .28$). In other words, changes in the leftmost grade is the primary source of a positive correlation between grade average and citizens’ evaluation of school performance.

In Figure 4 the leftmost-digit-bias is shown for the grades with the most observations (i.e., 5, 6, and 7). This

Figure 3: Linear regression lines for both the full range of the data and for data within the intervals of each of the leftmost grades.

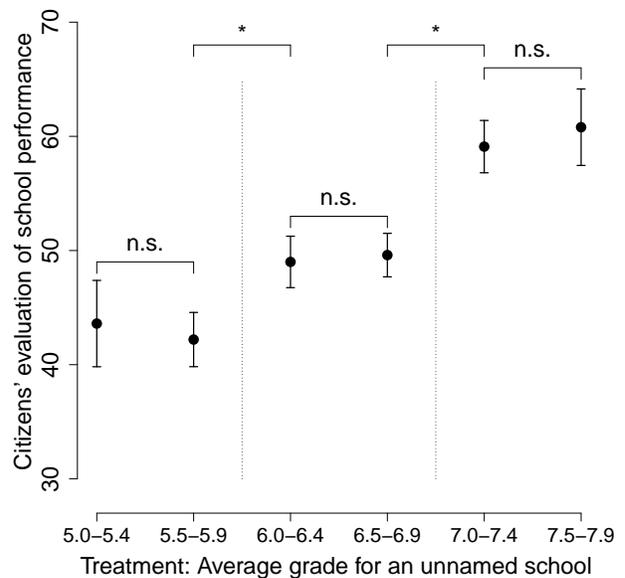


provides a more in-depth test of differences in evaluations within and between leftmost grades. Within each leftmost grade there is no significant difference in performance evaluation between schools with lower-end ($x.0-x.4$) and higher-end ($x.5-x.9$) decimals. However, as a school's grade reaches a shift in the leftmost grade digit, we observe substantial and significant shifts in citizens' performance evaluation. From the interval 5.5–5.9 to the interval 6.0–6.4 the average evaluation increases with 6.84 points ($t(339) = 3.92, p < .001$). The increased average evaluation from the interval 6.5–6.9 to the interval 7.0–7.4 is as high as 9.51 points ($t(396) = 5.92, p < .001$).

We can even observe these large effects in the very close vicinity of leftmost grade shifts. For instance, schools with a grade of 6.9 are rated significantly lower than schools with the slightly higher grade average of 7.0 (47.47 vs. 56.92 points, $d = 9.45, t(85) = 2.72, p < .01$). A similar pattern is found for the grade of 5.9 compared with 6.0. Here the tiny grade difference of .1 increased average evaluation by 7.28 points ($t(98) = 2.58, p < .05$). At the same time, there is no difference in evaluations between schools with 5.0 compared with 5.9 ($d = 4.48, t(69) = 1.03, p = .31$) or for those with 6.0 compared with 6.9 ($d = -2.66, t(92) = 0.86, p = .39$). On the one hand, a very small shift in the average grade which changes the leftmost grade can have large effects on evaluation. On the other hand, a shift 10 times larger leaves citizens' evaluation unaffected.

A simple overall analysis can summarize the effect. This is done using three simple predictors: the overall

Figure 4: Means with 95%–CIs for grade decimals in the intervals .0–.4 and .5–.9 for each of the leftmost grades 5, 6, and 7. A star indicates significance at $p < .001$.



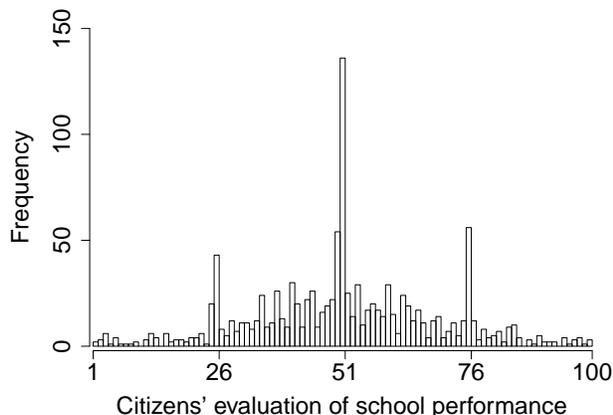
treatment grade, a continuous variable of the leftmost digits (3–9), and a continuous measure for the rightmost decimal (0–9). A regression of the evaluation on the leftmost digit and rightmost digit variables revealed a highly significant effect of leftmost digit (unstandardized coefficient 8.4, $p < .001$) and a non-significant effect of the rightmost digit (1.47, $p = .383$). And a regression of the evaluation on the full grade and the rightmost digits revealed a significant effect of the grade (8.4, $p < .001$) and a *negative* effect of the rightmost digit ($-6.93, p < .001$), showing that the effect of the decimals had to be removed from the full grade to get the best fit.

3.1 Robustness and additional analysis

Some further robustness checks and additional analysis should be noted briefly. These include: 1) accounting for clustering of answers on the response scale and outliers, and 2) test if numeracy (number skills) or familiarity with the scale mediates the degree of the leftmost-digit-bias. First, Figure 2 showed some graphical markers for the first quarter (26 points), the middle (51 points), and the last quarter (76 points). Figure 3 indicated a clustering of responses around these marks which is also visible in a simple histogram of the response variable as seen in Figure 5.

The analysis also indicated a few responses at the very extreme ends of the scale. It should therefore be checked how the graphically salient responses and potential outliers affect the results. In Figure 6 means for each bin

Figure 5: Citizens evaluation of an unnamed school’s performance given information about its grade average.



of the treatment grade is shown for the range of the data with the most observations and where the leftmost-digit-bias was found to be strongest. The regular means clearly show discontinuities in evaluations as the leftmost digit change. The results are very similar if trimmed means are applied. Finally, bin means excluding values 26, 51, and 76 do not seem to substantially alter the results.

The survey following the experiment provided some relevant information about the respondents. To ask whether socio-economic factors mediate the leftmost-digit-bias, I ran a number of simple regression analysis using dummies for leftmost digits and the continuous variable indicating the rightmost digit. I then tested the interaction of the rightmost decimal variable with age, gender and education of the respondents. These socio-economic factors capture potential differences in experience in numeracy. Furthermore, I tested similar interactions with two dummy variables; one for respondents with kids in school age and one for respondents with previous or current work experience from a school. These two dummies capture differences in familiarity with average grades. Around 10 years ago the educational system shifted the grade system from a 10-point-scale to the current 7-point-scale. The interactions between respondent characteristics and the measure of the rightmost decimal tests whether differences in familiarity or indicators of numeracy affect the results. No such interactions were significant: neither age, gender, education, kids in school, or working experience from a school moderate the extent to which rightmost-digits are discarded.

4 Conclusion and discussion

The analysis has shown that citizens’ evaluation of public sector performance given simple two-digit grade information is biased by a strong leftmost-digit-effect. The

substantial implication is that very small changes in the reported performance information, which happen to shift the leftmost digit, can lead to very large changes in citizens’ judgment of performance. Meanwhile, even large shifts in the average performance within the same leftmost digit have no significant effects on judgment about performance. Public opinion about public sector performance is therefore likely to follow a stepwise function formed by leftmost digits.

The hypothetical setup of the experiment can limit the extent to which the findings are applicable in real-world political settings. This being said, the results provide a reasonable explanation for why studies have found that reelection-minded politicians arrange numerical policy information in favorable ways with regards to leftmost digits (Brunell & Glazer, 2001; Ashworth, Heyndels, & Smolders, 2003; Olsen, 2013). These experimental results are therefore coherent with descriptive findings from real-world political settings.

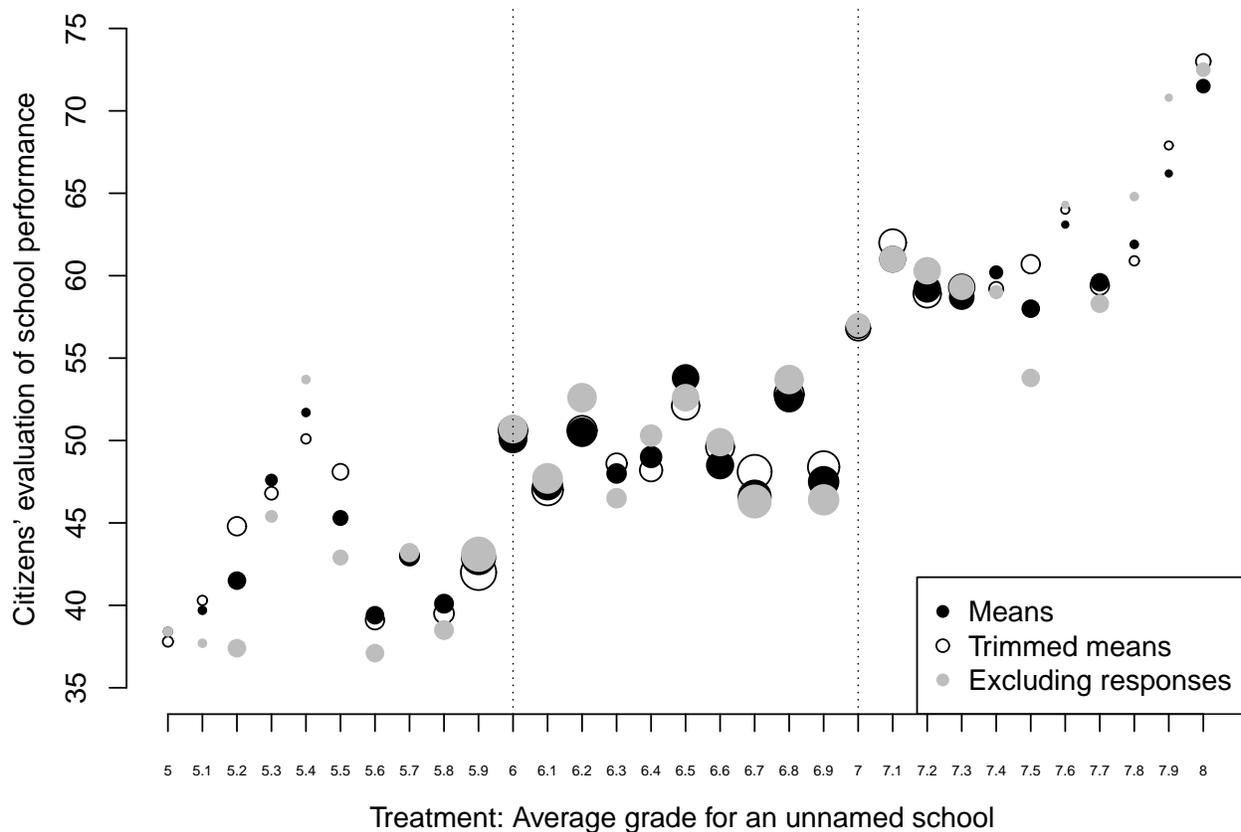
The results also extend and confirm the experimental and observational work on the leftmost-digit-bias in market settings (Schindler & Kibarian, 1996; Lacetera, Pope, & Sydnor, 2012). The particular case of school grades denote that the bias is present even for familiar metrics of a very simple two-digit form.

The results extend findings in recent studies showing that invisible taxes can shift behavior when they are made visible, e.g., by posting supermarket prices with sales tax included on the price tag (Chetty, Looney, & Kroft, 2009; DellaVigna, 2009; Finkelstein, 2009). With the leftmost-digit-bias, numbers can be processed in a biased way while still being fully visible.

Future research must look at the conditions under which biases in human processing of numerical performance information is reduced or enhanced. One possibility is to see how the introduction of multiple reference points change the leftmost-digit-bias. For instance, the leftmost-digit-bias can be seen in context of studies on “category boundaries” (Rothbart & Davis-Stitt, 1997) where we can think of leftmost-digits as categories which form a boundary for any digit further to the right. The question is how the bias is affected if other categorical boundaries are made salient. For instance, certain labels or thresholds can be assigned to ranges of numerical information. The present results do not lend much insight into this question. However, there is some indirect evidence that the leftmost-digit-bias can co-exist with other categorical boundaries. The largest leftmost-digits-bias was identified between the grade shifts of 5–6 and 6–7. Of these grade averages only 7 is an actual grade which has its own labeling as “slightly above average”.² That

²A translation from the grading description reads: “The grade 7 is given for a good performance displaying good command of the relevant material, but also some weaknesses.” 7 is seen as equivalent to the

Figure 6: Robustness checks of the findings in the grade interval 5.0–8.0. Dots represent means for each bin of the treatment grade. Dot sizes reflect the number of observations in each bin. Larger dots indicate more observations. In the displayed interval the number of observations in each bin of the regular means varies from a minimum of 13 to a maximum of 53.



is, grades “5” and “6” exist only as grade averages but not as formal grades. Yet the analysis showed similar sized discontinuities in evaluations between 5–6 as seen between 6–7. This indicates that grade “7”’s status as a “real” grade did not enhance it’s leftmost digit influence.

A further understanding of citizens’ biased number processing is of great importance in an increasingly “enumerated” public sector. On the one hand, performance information provides important comparable data to inform the judgments and choices made by citizens. On the other hand, if even simple two-digit information is evaluated in a biased way, performance information is likely to introduce new challenges to the task of holding politicians and public service providers accountable.

References

Anderson, E. T., & Simester, D. I. (2003). Effects of \$9 price endings on retail sales: Evidence from field experiments. *Quantitative Marketing and Economics*, 1, 93–110.

Ashworth, J., Heyndels, B., & Smolders, C. (2003). Psychological taxing in Flemish municipalities. *Journal of Economic Psychology*, 24, 741–762.

Bizer, G. Y., & Schindler, R. M. (2005). Direct evidence of ending-digit drop-off in price information processing. *Psychology and Marketing*, 22, 771–783.

Bowman, N. A., & Bastedo, M. N. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of US News and World Report on student decisions. *Research in Higher Education*, 50, 415–436.

Brunell, T. L., & Glazer, A. (2001). Rational response to irrational attitudes: The level of the gasoline tax in the United States. *Journal of Policy Analysis and Management*, 20, 761–764.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99, 1145–77.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regu-

American “C”.

- larities in the frequency of number words. *Cognition*, 43, 1–29.
- DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47, 315–372.
- Dixon, R., Hood, C., & Jones, L. R. (2008). Ratings and rankings of public service performance: Special issue introduction. *International Public Management Journal*, 11, 253–255.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 133, 1–40.
- Finkelstein, A. (2009). EZ-Tax: Tax salience and tax rates. *Quarterly Journal of Economics*, 124, 969–1010.
- Hinrichs, J. V., Berie, J. L., & Mosel, M. K. (1982). Place information in multidigit number comparison. *Memory and Cognition*, 10, 487–495.
- Korvost, M., & Damian, M. F. (2008). The differential influence of decades and units on multidigit comparison. *Quarterly Journal of Experimental Psychology*, 61, 1250–64.
- Krishna, A., & Slemrod, J. (2003). Behavioral public finance: Tax design as price presentation. *International Tax and Public Finance*, 10, 189–203.
- Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, 102, 2206–2236.
- McCaffery, E. J. (1994). The UCLA tax policy conference: Cognitive theory and tax. *UCLA Law Review*, 41, 1861–1947.
- McCaffery, E. J., & Baron, J. (2006). Thinking about tax. *Psychology, Public Policy, and Law*, 12, 106–135.
- Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the US News and World Report college rankings. *Research in Higher Education*, 45, 443–461.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, 63, 81–97.
- Olsen, A. L. (2013). The politics of digits: Evidence of odd taxation. *Public Choice*, 154, 59–73.
- Peters E., Slovic, P., Västfjäll, D., & Mertz C. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making*, 8, 619–635.
- Pollock, S. E., & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology*, 10, 32–45.
- Pope, D. G. (2009). Reacting to rankings: Evidence from “America’s Best Hospitals”. *Journal of Health Economics*, 28, 1154–65.
- Pope, D. G., & Simonsohn, U. (2011). Round numbers as goals. *Psychological Science*, 22, 71–79.
- Prewitt, K. (1987). Public statistics and democratic politics. In Alonso, W., & Starr, P. (Eds.), *The politics of numbers*. New York: Russell Sage Foundation Publications (pp. 261–274).
- Rothbart, M., & Davis-Stitt, C. (1997). Effects of arbitrarily placed category boundaries on similarity judgment. *Journal of Experimental Social Psychology*, 33, 122–145.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532–547.
- Schindler, R. M., & Kibarian, T. M. (1996). Increased consumer sales response through use of 99-ending prices. *Journal of Retailing*, 72, 187–99.
- Schindler, R. M., & Kirby, P. N. (1997). Patterns of rightmost digits used in advertised prices: Implications for nine-ending effects. *The Journal of Consumer Research*, 24, 192–201.
- Smith, A. R., & Windschitl, P. D. (2011). Biased calculations: Numeric anchors influence answers to math questions. *Judgment and Decision Making*, 6, 139–146.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277–310.
- Thomas, M., & Morwitz, V. (2005). Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research: An Interdisciplinary Quarterly*, 32, 54–64.