Benford's Law: Textbook Exercises and Multiple-Choice Testbanks

Aaron D. Slepkov[1]*, Kevin B. Ironside[1], David DiBattista[2]

1 Department of Physics & Astronomy, Trent University, Peterborough, Ontario, Canada, **2** Department of Psychology, Brock University, St. Catharines, Ontario, Canada

* aaronslepkov@trentu.ca

Abstract

Benford's Law describes the finding that the distribution of leading (or leftmost) digits of innumerable datasets follows a well-defined logarithmic trend, rather than an intuitive uniformity. In practice this means that the most common leading digit is 1, with an expected frequency of 30.1%, and the least common is 9, with an expected frequency of 4.6%. Currently, the most common application of Benford's Law is in detecting number invention and tampering such as found in accounting-, tax-, and voter-fraud. We demonstrate that answers to end-of-chapter exercises in physics and chemistry textbooks conform to Benford's Law. Subsequently, we investigate whether this fact can be used to gain advantage over random guessing in multiple-choice tests, and find that while testbank answers in introductory physics closely conform to Benford's Law, the testbank is nonetheless secure against such a Benford's attack for banal reasons.

Introduction

The expectation that the leading digits in the numbers one encounters in everyday life are uniformly distributed is a well-known "pitfall of elementary statistics" [1]. As early as 1881, mathematician Simon Newcomb observed that the earlier pages in his bound logarithm tables were considerably more worn than the later pages in the book [2]. This led him to posit that the leading digits of various numbers he encountered were distributed logarithmically. Fifty years later, physicist Frank Benford independently observed the same effect in his log tables and came to the same conclusion. Benford then went on to collect over 20,000 numbers from 20 different datasets such as front-page newspaper entries, street addresses of eminent scientists, physical constants, and American League baseball statistics, to show that this finding applies to a wide range of phenomena [3]. Benford's proposed distribution of leading digit frequencies is given by

$$P_i^B = \log_{10}\left(\frac{i+1}{i}\right); \quad i \in \{1, 2, 3, ..., 9\}, \tag{1}$$

where P_i^B is the probability of finding i as the leading digit in a given number. Therefore, the frequency of leading digits diminishes monotonically from 30.1% and 17.6% for the

digits 1 and 2, to 5.1% and 4.6% for the digits 8 and 9, respectively. Benford called this finding a "law of anomalous numbers" [3]. Nonetheless, in accordance with Stigler's Law of eponymy [4], this has come to be known as Benford's Law of leading digits.

Must there be a well-defined distribution of leading digits for a given phenomenon or dataset? Intuitively, most people suspect that the most likely distribution of leading digits should be a uniform distribution, such that they are all equally likely. However, for a given physical phenomenon, any such stable distribution should be scale-invariant in a way that the choice of unit system or choice of base would maintain the existence of such a distribution. For example, if there is to be a well-defined distribution of leading digits for a tabulation of the mass of various insect species, it shouldn't matter if the mass is measured in milligrams or in grains of rice. It has been shown that Benford's Law gives the only scale invariant distribution of leading digits [5]. Furthermore, such scale invariance assures base invariance as well [6]. Thus, it could be said that if a given dataset is to have an identifiable leading-digit distribution, then it must follow Benford's Law [7]. However, there is no requirement that all datasets must have a base-invariant distribution of leading digits, and thus not all phenomena follow Benford's Law.

Decades of research have led to some guidelines for predicting which datasets might follow Benford's Law [8–10]. Such phenomena should span several orders of magnitude, and preferably should be unbounded. The numbers in the dataset should be physically-relevant and should represent a measurement that has associated units; phone numbers, lottery numbers, and license plate numbers are not expected to be Benford-distributed. Finally, it has been suggested that the dataset should contain more small quantities than large quantities [8, 11]. Despite such guidelines, there are numerous examples of sets of numbers that violate at least one of these guidelines and still follow Benford's law, such as the Fibonacci Numbers [12], and most geometric series [3]. There are now numerous published examples of mathematical series and physical datasets that follow Benford's Law [13]. These include geological streamflow rates [14], radioactive half-lives [15], hadron widths [16], numerous statistical distributions [17, 18], auction prices [19], and business invoices and tax returns [8, 9, 20], as well as the original datasets by studied Benford [3]. The precision with which various types of data follow Benford's Law has led to its widespread application for forensic accounting and auditing [8, 21, 22], and while this application represents by far the most active (and commercial) use of the law, there are nonetheless serious practical concerns regarding the validity and justification of this practice [9, 23]. Benford's Law has not yet found robust application in the sciences, although from early on there have been suggestions that it could be used for diagnostic applications in computer design and scientific calculation errors [5, 6]. Indeed, there has often been a sense that Benford's Law is purely mathematical, rather than physical, and will never find practical applications [24]. Others argue that the law is wholly physical, and thus has tremendous potential for future applications [9].

Because only certain datasets that correspond to real phenomena closely follow Benford's Law, we trust that knowledge of this fact could be applied towards pro-active or predictive ends—as opposed to *a posteriori* diagnostics. In consideration of the general guidelines for datasets that are expected to conform to Benford's Law, we identify the set of answers to end-of-chapter questions in introductory physics and chemistry textbooks as a possible Benford's Law candidate. In fact, prior research has described the relative abundance of small-value numbers in grade school math textbooks that focus on arithmetic and multiplication, and has suggested a possible link to Benford's Law [25, 26]. In this article, we test this hypothesis, and find the set of introductory science textbook answers to conform well to the logarithmic distribution of leading digits.

Confirmation that quantitative answers to a wide range of "problems" in physics and chemistry follow Benford's Law immediately suggests a practical (if somewhat nefarious)

application: examination questions that are meant to assess knowledge of introductory physics and chemistry should be similar to those found in the practice sections of the textbooks. Thus, multiple-choice testbanks that are created to provide a valid examination tool might also follow similar leading-digit trends found in the textbook exercises. If this is true, then perhaps the implication that over 60% of the answers to numerical multiple-choice exam questions are anticipated to have a leading digit of 1, 2, or 3 can be utilized to gain an advantage by physics-ignorant but test-wise students. Traditionally, a physics-ignorant student can resort to complete guessing on a multiple-choice test, and thus each question has an expected baseline for guessing that is based on the number of options in the question. A student has a 33% chance of getting any given question right in a 3-option question, a 25% chance in a 4-option question, and a 20% chance in a 5-option question. Students might employ a slew of test-wise strategies to boost this baseline [27]. One such strategy—which is often derided by professors but may have some marginal advantage [28]—involves preferential selection of middle options. This manifestation of edge-avoidance is colloquially known as "when in doubt, choose C!". This strategy (as well as most others) is based on poor test construction, and is easily foiled. On the other hand, a Benford's Law-based attack on a multiple-choice testbank would be guided by the assumption that it is more likely that the incorrect options (distractors) are semi-randomly selected in such a way that each yields a uniform first-digit distribution. Thus, after demonstrating that answers to end-of-chapter textbook questions yield a Benford distribution, we discuss considerations for which a Benford's Law-based attack on a multiple choice testbank is expected to gain an advantage over random guessing. We then proceed to analyze the distribution of leading digits in an actual introductory physics multiple-choice testbank, and go on to find that when we attack the testbank directly the ubiquity of the Benford distribution in fact secures the bank against such attacks.

Finally, we discuss how the rounding off of numbers to a pedagogically-motivated reduced set of significant figures is expected to alter the distribution of leading digits away from Benford's Law and demonstrate that a modified distribution based on this fact yields even better fits to testbank data in physics.

Methods and Results

Data Collection

To select an appropriate representative sample of physics and chemistry exercises, three books were chosen based on ready availability on our bookshelf and on current popularity in undergraduate physics education:

- "Physics for Scientists and Engineers: A Strategic Approach", 3rd edition, by Knight (Pearson Education, 2013) is a highly popular introductory physics textbook with an approach that is based on recent physics education research findings.

- "Sears and Zemansky's University Physics", 10th edition, by Young & Freedman (Addison Wesley Longman, 2000) was a popular calculus-based introductory physics textbook a decade ago, with emphasis on physics education fundamentals.

- "Fundamentals of Analytical Chemistry", 7th edition, by Skoog, West, and Holler (Saunders College Publishing, 1996) was a favorite intermediate-level undergraduate chemistry textbook for one of us (ADS), and was selected in anticipation of useful quantitative chemistry end-of-chapter questions.

Henceforth these three books will be referred to as Knight, Young & Freedman, and Skoog, respectively. The data was obtained by recording the leading digit (i.e. the leftmost nonzero digit)

Table 1. Obtained frequencies of leading digits in textbooks and testbank.

		Benford dist.	Knight 3rd ed. end-of-chapter answers	Young 10th ed. end-of-chapter answers	Skoog 7th ed. end-of-chapter answers	Knight testbank answers	Knight testbank distractors	combined data
# entries			1644	2155	294	1485	5671	11249
	1	0.3010	0.2889	0.2858	0.3129	0.2774	0.2906	0.2883
	2	0.1760	0.1673	0.1689	0.1667	0.1872	0.1816	0.1774
	3	0.1249	0.1247	0.1128	0.1224	0.1428	0.1259	0.1253
	4	0.0969	0.0991	0.1104	0.1122	0.0875	0.1010	0.1011
i	5	0.0791	0.0803	0.0891	0.0714	0.0714	0.0890	0.0850
	6	0.0669	0.0803	0.0733	0.0646	0.0835	0.0615	0.0695
	7	0.0579	0.0566	0.0617	0.0510	0.0606	0.0531	0.0562
	8	0.0511	0.0535	0.0599	0.0340	0.0505	0.0474	0.0508
	9	0.0457	0.0493	0.0381	0.0646	0.0391	0.0497	0.0464
MAD[1]			0.0050	0.0094	0.010	0.011	0.0054	0.0033
			close conform	accept. conform	accept. conform	accept. conform	close conform	close conform
SSD[2]			4	9	12	15	3	2
			accept. conform	accept. conform	accept. conform	accept. conform	accept. conform	perfect conform

[1] MAD = Mean Absolute Deviation (see text for definition)

[2] SSD = Sum of Square Differences (see text for definition)Obtained distributions of leading digits and measures of conformation to Benford's Law for three textbooks, a multiple-choice testbank, and aggregate data

from every end-of-chapter answer. Data collection was implemented manually by parsing through the texts, but with a protocol designed to eliminate subjective selection. Nonetheless, because of the general constraints of anticipated Benford data sets, such as avoiding unphysical numbers and numbers that are too narrowly confined in domain, we rejected all unitless values, values reported as percentages, and those with units of degrees. Furthermore, because the number zero is meaningless from a significant digit standpoint, all answers of exactly zero were rejected. Obviously, non-numeric end-of-chapter answers such as pictures, graphs, equations, and textual answers were ignored. In all, approximately 10%–15% of entries in the introductory physics texts were rejected and 30%–35% were ignored as non-numeric. In Skoog—the analytical chemistry text—as many as 25% of the numeric answers were rejected, largely due to being unitless or percentages. Other than these limitations, all other data was recorded. For testbank data, we recorded leading digits of the multiple-choice numerical items supplied with Knight. When parsing this testbank we looked at both the keyed options and at the distractors, recording the leading digit of each separately. In recording testbank entries, the same protocol was followed as described for textbook data; however, the number of rejected entries in the testbank was only 7%. In sum, this data is presented in Table 1. The full raw dataset is freely available as S1 Dataset.

Data Analysis

The data presented in Table 1 includes the theoretically-expected leading digit distributions of Benford's Law. According to Eqn. 1 this distribution can never be perfectly realized in any dataset because the values are irrational numbers. Thus, Benford's Law can only be *approached*, and any dataset—no matter how good—will ultimately deviate from the ideal distribution. In many cases, where the invocation of Benford's Law is meant simply to highlight the
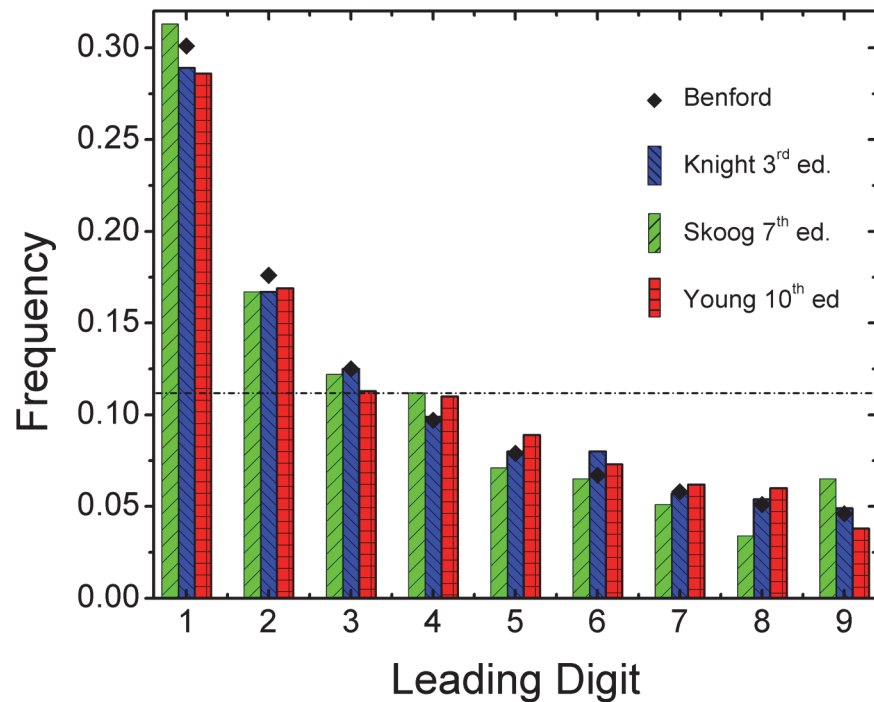
Fig 1. The distribution of leading digits in end-of-chapter excercise answers from two popular introductory physics textbooks (Knight, Young & Freedman) and an analytical chemistry textbook (Skoog). The dashed horizontal line indicates uniform distribution of first digits. Overlaid black squares are the theoretical Benford's Law distribution.

disproportionate abundance of low-value leading digits over high-value leading digits, the suggestion of Benford's Law can be confirmed by visual inspection of a digit frequency histogram. Such a histogram is presented in Fig. 1, showing the distributions for end-of-chapter exercise answers from the three textbooks. All three textbooks clearly yield a Benford-like distribution with an observed monotonic decrease in frequency with increasing leading digit value. Establishing a benchmark statistical measure of conformity to Benford's Law has been an ongoing research endeavor [8, 9, 19, 29]. In the Benford's Law literature within the natural sciences, the most common measure of statistical conformity to a Benford distribution is the χ^2 test for goodness of fit. However, leading expertise in Benford's Law analysis finds that the χ^2 test is widely misused and misinterpreted [9]. The χ^2 test has been described as suffering from an "excess power" problem, wherein larger data sets require increasingly better fits to pass the χ^2 threshold for conformity [8]. Thus, larger data sets that by inspection give better fits than smaller datasets will often fail a χ^2 test that the smaller dataset passes [30].

As a way to avoid problems with the χ^2 test, other simple tests have been championed, as a means of judging dataset conformity to Benford's Law. For example, Nigrini has proposed using a mean absolute deviation (MAD) measure [8], and Kossovsky has proposed using a sum of squares difference (SSD) measure [9]. MAD is an empirically-based whole-test measure that simply takes the average of the absolute deviation of each digit's frequency from the ideal Benford's Law frequency. Specifically, this is given by

$$\text{MAD} = \frac{\sum_{i=1}^{K} |AP - EP|}{K},$$

(2)

where K is the number of leading digit bins (9 for first leading digit; 90 for first two leading digits, etc.), AP is the actual proportion observed, and EP is the expected proportion according to Benford's Law. The MAD test does not have an analytically-derived critical value. Instead, Nigrini has established empirically-based criteria for conformity to Benford's Law. [8] The suggested MAD ranges for "close conformity", "acceptable conformity", and "marginal conformity" are 0.000–0.006, 0.006–0.012, and 0.012–0.015, respectively. A MAD value above 0.015 is considered non-conforming.

Likewise, SSD too is an empirically-based whole-test measure that simply takes the sum of the squares of the deviation of each digit's frequency from the ideal Benford's Law frequency. Specifically, this is given by

$$\text{SSD} = \sum_{i=1}^{K} (AP - EP)^2 \times 10^4, \tag{3}$$

Like the MAD test, the SSD test does not have an analytically-derived critical value. Kossovsky's suggested SSD conformity criteria for "perfect conformity", "acceptable conformity", and "marginal conformity" are 0–2, 2–25, and 25–100, respectively. An SSD value above 100 is considered non-conforming [9].

Results

The three textbooks varied in the number of available entries for analysis, but overall contained a sufficient number of entries and domain of data to warrant a Benford's Law analysis. Skoog provided the smallest data set with a total of 294 entries in 32 chapters, with several chapters having no usable entries. In total, however, the numerical entries in this book spanned 60 orders of magnitude. Knight and Young both provided large data sets, with 1644 and 2155 entries in 42 and 39 chapters, respectively. Numerical answers to physics questions in these books spanned over 170 orders of magnitude. As seen from Table 1, Knight closely conforms to Benford's law (MAD = 0.0050; SSD = 4), while Young & Freedman (MAD = 0.0094; SSD = 9) and Skoog (MAD = 0.010; SSD = 12) both show acceptable conformity. A visual inspection of the distribution histogram presented in Fig. 1 also strongly suggests close conformity. Thus we conclude that, in general, numeric end-of-chapter questions in physics and chemistry textbooks follow Benford's Law.

The first-digit frequency distribution for the companion multiple-choice testbank to Knight is presented in Fig. 2. The keyed-responses (i.e. correct answers) conform acceptably to Benford's Law, yielding a MAD value of 0.011 and an SSD value of 15. Thus, over 50% of the answers to numerical multiple-choice exam questions are anticipated to have a leading digit of 1, 2, or 3. As mentioned above, a Benford's Law-based attack on a multiple-choice testbank would be guided by the assumption that it is more likely that the incorrect options (distractors) are semi-randomly selected in such a way that each yields a uniform first-digit distribution. Then the question remains whether the latent Benford distribution of the keyed response will provide more low-digit options than an ensemble of uniformly distributed distractors. The frequency of each *lowest*-leading-digit, i, among a group of N distractors can be given by

$$P_N^i = \frac{1}{9^N} \sum_{k=1}^{N} (9-i)^{N-k} \binom{N}{k}, \tag{4}$$

where $\binom{N}{k}$ are the standard binomial coefficients (often read as "N choose k"). These probabilities are presented in Fig. 3, for each of a 3-, 4-, and 5- option multiple choice test, overlaid with the Benford-distributed correct response.
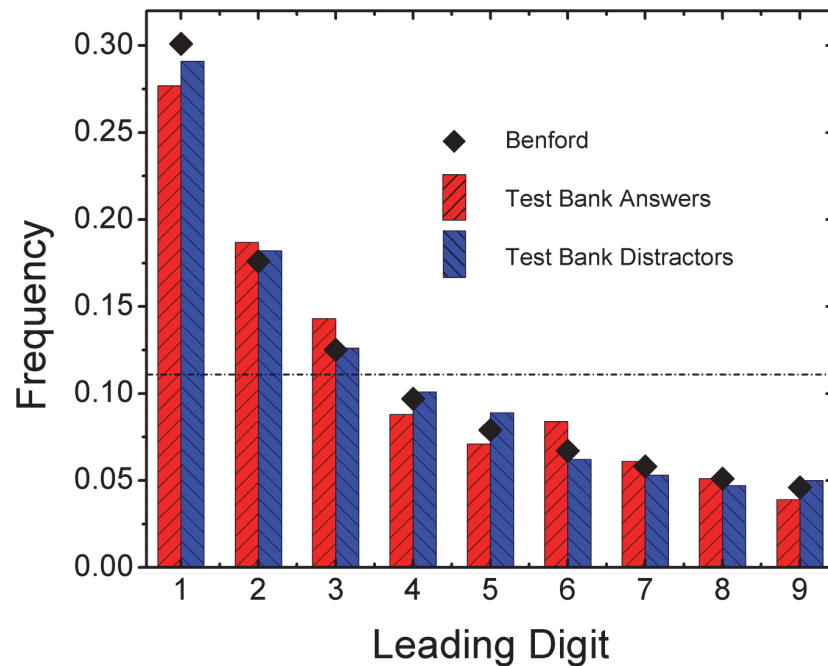
Fig 2. The distribution of leading digits in multiple-choice testbank answers and associated distractors for Knight, "Physics for Scientists and Engineers", 3rd edition. The dashed horizontal line indicates uniform distribution of first digits. Statistical analysis confirms conformation to Benford's Law, overlaid as black squares. Conformation of the distractors to Benford's Law precludes a Benford-based attack on the testbank.

doi:10.1371/journal.pone.0117972.g002

From Fig. 3, we see that, while for a 3-option test the lowest leading digit 1 is more probable in the Benford-distributed keyed responses than in the combined pair of two distractors, at least one distractor in each 4- and 5-option item is expected to have *on average* a lower first digit than the keyed response. These considerations would strongly moderate the advantage of a Benford's Law-based attack on a set of multiple-choice questions, but would not entirely ne-gate this strategy. This is demonstrated by considering the probability of finding the correct an-swer by choosing the lowest leading digit among a group of N uniformly-distributed-first-digit-distractors and a Benford's-distributed correct answer, with random guessing in cases where the lowest leading digit is not unique, as given by

$$P_N^B = \sum_{i=1}^{9} \frac{P_i^B}{9^N} \left[\sum_{k=0}^{N} (9-i)^{N-k} \frac{1}{k+1} \binom{N}{k} \right], \qquad (5)$$

where P_i^B are the Benford probabilities as given by Eq. 1. This expression is analogous to that presented in a recent note by F. M. Hoppe [31], and is plotted for 2- to 6-option multiple choice questions in the inset to Fig. 3. Using Eq. 5 we find that such an attack always improves the test score over random guessing: For a 3-, 4-, or 5-option test we expect scores of 53%, 44%, or 38%, respectively. Compared to blind-guessing scores of 33%, 25%, and 20%, a Benford attack promises a significant advantage. The expectation of a passing score in a 3-option exam is par-ticularly noteworthy considering expert recommendations that this is (psychometrically) the most desirable type of multiple-choice item [32].

We attacked the Knight testbank by selecting the option with the lowest leading digit or guessing among items with identical lowest leading digits, and compared these selections to the
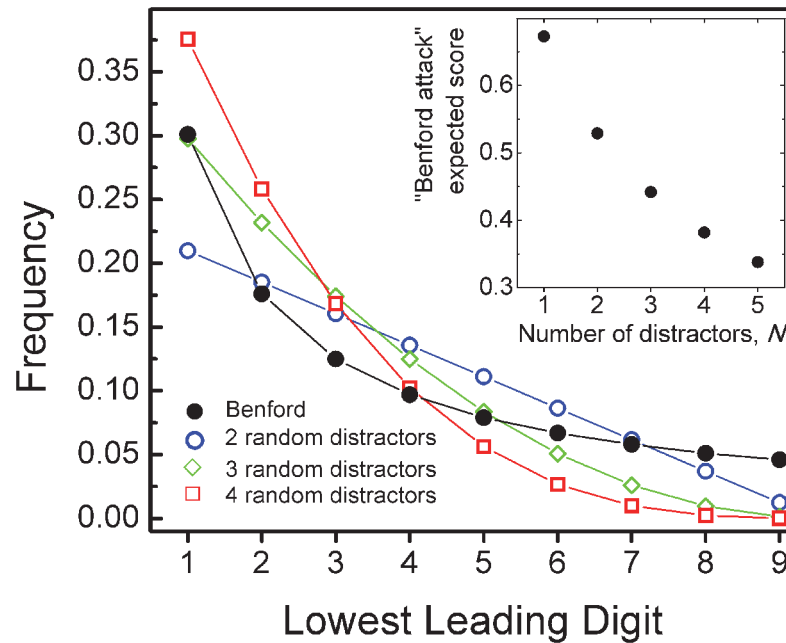
Fig 3. The frequency of lowest-leading-digits among an ensemble of uniformly-distributed-fist-digit distractors _N_ for 3-, 4-, and 5-option multiple choice questions, and the expectation of success using a Benford's attack on such a test. For a Benford's Law-based attack on a testbank the predominance of low-value leading digits in the answers must be maintained in the presence of a group of distracotrs. Despite the fact that for 4- and 5-option questions the distractors are collectively more likely to have the lowest leading digit, a Benford attack on such a group is nonetheless expected to yield an advantage over a random-guessing strategy (inset). In the case of a test with 3-option questions—where the answers are Benford distributed and the two distractors are uniformly distributed—a Benford attack is expected to yield a passing score of 53% (inset).

doi:10.1371/journal.pone.0117972.g003

keyed response. The majority of questions were of the 4-option type. Our attack yielded a score of 24.6%; i.e. no better than chance. The explanation for the failure in this strategy lies in the first-digit distribution of the distractors. As shown in Fig. 2 (and listed in Table 1), the testbank distractors also closely conform to Benford's Law, meeting the strictest MAD criteria. Thus, the fact that the keyed responses are Benford distributed is marginalized by the likewise distributed distractors.

Discussion

As we have shown, typical physics and chemistry questions, as a group, follow Benford's Law for leading digits, as do both the keyed responses and the distractors of a large introductory physics testbank. Uniformly-distributed random numbers, however, do not follow Benford's Law, but rather have uniformly-distributed leading digits. Thus, one definite conclusion we can draw from our analysis of the testbank is that for this set of multiple-choice questions the distractors are clearly not random numbers. A more interesting question is whether the fact that the distractors follow the same pattern of leading digits as do the keyed-responses—which emerge from well-defined and deterministic procedures—means that they too are a result of a similar creation process? That is, are the distractors necessarily created as answers to alternate questions? Not necessarily. While *de novo* random numbers are uniformly distributed in leading-digit, processed random numbers are not, and multiplications of arrays of random number are known to generate near-ideal Benford sets [33]. Furthermore, a random selection of

numbers from multiple sets of different distributions—none of which need to be Benford distributed—yields a Benford set. This appears to be a leading-digits analogue to the central limit theorem [3, 6]. Finally, the scale invariance of Benford's Law suggests that multiplying values within a Benford set by various other numbers maintains the distribution of first digits. Thus, we can not identify the way in which the distractors were created. Whether they are made of processed random numbers, processed from the keyed response, or are themselves erroneous but plausible answers to a set of questions cannot be discerned from their distribution of leading digits; any of these would yield a Benford Distribution.

The question of why a large but seemingly random subset of all possible quantitative physics questions should follow Benford's Law with such precision is warranted. There may not be a clear-cut answer to this question, but the truth probably lies in the aforementioned theorem that random sampling from a wide mixture of first-digit distribution sets converges to a Benford distribution. As a group, end-of-chapter questions (or potential final examination questions) span many topics and involve numerous different parameters, each of which may have a different first-digit distribution within the domain of physically relevant phenomena. As examples, the sets of likely "kinetic energies for vehicles on earth" and realistic "currents induced in copper rings by Lenz's Law" may be sufficiently limited in domain as to individually refrain from a Benford distribution, but random sampling of values across such distributions will yield a Benford set. This is likely the reason that such a good Benford distribution is found in our samples. As further evidence of this argument, we see that combining the data from the three textbooks, the testbank answers, and the testbank distractors yields an aggregate data set that better conforms to Benford's Law than any of the individual data sets. Included in Table 1, this combined dataset shows a minuscule MAD of 0.0033 and a SSD criteria of "perfect Benford".

Thus far we have determined that answers to physics questions conform to Benford's Law. However, these numerical quantities, as found in the textbooks and in the testbank, are often reported to an artificially reduced number of significant digits. The link between rounding and Benford's Law was identified at inception, [3] and has formed the basis for some potential applications of the law for computer design [5, 34]. Typical datasets reported in the Benford's Law literature contain many more than three significant digits. For pedagogical reasons, numbers reported in introductory physics courses are limited to the smallest number of significant digits warranted by the precision of the values used for the calculation or measurement. Instructional material with numbers that are reported to two or three significant figures may thus represent an artificial data set with an imperfect Bendord distribution of leading digits. A description of how rounding modifies the distribution of leading digits is fairly straightforward: In the case of first-digit frequencies, the effect of rounding is to diminish the expected number of 1s and increase the frequency of all other digits. We present a modification to Eq.1 that includes the effects of rounding, where N_{SD} is the number of significant digits to which a value is rounded, and $N_{SD} = \infty$ represents the traditional case of no rounding:

$$
P_i^B(N_{\text{SD}}) = \begin{cases} \log\left(\dfrac{i + 1 - \frac{1}{2}10^{-(N_{\text{SD}}-1)}}{i - \frac{1}{2}10^{-(N_{\text{SD}})}}\right) & \\ & i = 1 \\ \log\left(\dfrac{i + 1 - \frac{1}{2}10^{-(N_{\text{SD}}-1)}}{i - \frac{1}{2}10^{-(N_{\text{SD}}-1)}}\right) & \\ & i \in \{2, 3, ..., 9\} \end{cases} \tag{6}
$$

Table 2. Effects of rounding on Benford distribution.

D_i	Benford's Law $N_{SD} = \infty$	$N_{SD} = 2$	$N_{SD} = 1$
1	0.301	0.292	0.198
2	0.176	0.180	0.222
3	0.125	0.127	0.146
4	0.0969	0.0980	0.109
5	0.0792	0.0799	0.0872
6	0.0669	0.0675	0.0726
7	0.0580	0.0584	0.0622
8	0.0511	0.0515	0.0544
9	0.0458	0.0460	0.0483
MAD		0.0020	0.023
SSD		1	134
		close conform	non-conform

Comparing Benford distributions for leading digit in datasets with numbers rounded to one and two significant digits

doi:10.1371/journal.pone.0117972.t002

For numbers rounded to three or more significant digits the distribution of the leading digit is nearly identical to Benford's Law. However, in the case of values rounded to two significant digits, there is a small but significant modification of the first few digits. If numbers are rounded to only one significant digit, the distribution of this leading digit varies drastically from Benford's Law to the point where the occurrence of the digit 2 becomes more probable than the digit 1. Table 2 summarizes this relationship, showing that a dataset comprised of values rounded to two significant digits would perfectly pass both the MAD and SSD tests of conformity to Benford's law, but a dataset of values rounded to a single significant digit would fail to conform to a Benford's Law test despite the conformity of the underlying phenomena (i.e. the un-rounded data). We have observed that approximately 30% of entries in the testbank are reported to two significant digits. Thus, perhaps for the testbank a more appropriate distribution of leading digits than that given by Eq.1 should be given by a respective 0.7:0.3 weighted average of Eqns.1 and 6. When compared to this hybrid distribution, we obtain a MAD value of 0.0040, which is slightly better than the MAD value of 0.0045 of this full testbank dataset to the unmodified Benford's Law. Likewise, the SSD reduces from 3 to 2, markedly improving the data's fit to the hybrid distribution. Thus, we are likely observing the effects of rounding in our data, and this effect is expected to be a factor in the leading-digit analysis of any similar dataset where rounding to one or two significant figures is common practice.

Summary and Conclusions

We recorded the leftmost significant digit of the answers to every end-of-chapter question in two popular introductory physics textbooks, an intermediate analytical chemistry textbook, and a large introductory physics multiple-choice testbank, and find that all conform to Benford's Law. The fact that the answers to multiple-choice testbank questions follow this trend suggested a means by which the testbank could be attacked by a subject-ignorant but test-wise student. We find that among a set of distractors (wrong answers), each having uniformly distributed leading digits, the Benford distribution of the keyed response could be used to pass a 3-option test. Nonetheless, when using this information to "guess" the correct answers in a real testbank we find that the distractors themselves conform to Benford's Law, thereby securing

the testbank from such an attack. The fact that the distractors are Benford distributed does not, however, assure that they were engineered intelligently, but simply suggests that they do not comprise uniformly-distributed random values. Thus, we expect any such testbank to be equally secure against a Benford's Law based attack. We observed that physics textbooks and testbank items are often reported rounded to two significant digits, and we have shown how this fact may have impacted its distribution of leading digits, modifying it slightly from Benford's law and primarily yielding a relative dearth of leading 1s. We expect that for many readers our demonstration of Benford's Law in end-of-chapter textbook answers will prove counter-intuitive, just as math educators were surprised that grade school texts were biased toward small addition and multiplication facts [25, 26]: most people tend to believe that the leading digits of random values, and the numbers in arithmetic problems, would be uniformly distributed. Armed with the knowledge that answers to physics questions follow Benford's law, one trifling piece of advice we can give to the test-wise student is as follows: at the end of a long constructed-response examination, if you have little time to double-check the answers to all of the questions, spend time on those questions that yielded final answers that have the largest leading digits; questions are expected to have answers with leading digits 7, 8, or 9 only 15% of the time. You should note, however, that any advantages of using Benfords Law will only pay dividends if working with a large data set, and thus any advantages in test-taking would be, at most, marginal.

Supporting Information

S1 Dataset. The full raw data and example statistical analysis file. Worksheet "Data" includes raw data and summary counts/proportions collected from the five principle sources described in the article. Worksheet "Stat Analysis" provides an example of the data analysis preformed on one of the data sources from "Data".
(XLS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: ADS DD. Performed the experiments: KBI ADS. Analyzed the data: KBI ADS. Contributed reagents/materials/analysis tools: ADS DD. Wrote the paper: ADS DD.

References

1. Goudsmit SA, Furry WH (1944) Significant figures of numbers in statistical tables. Nature 154: 800–801. doi: 10.1038/154800a0

2. Newcomb S (1881) Note on the frequency of use of the different digits in natural numbers. Am J Math 4: 39–40. doi: 10.2307/2369148

3. Benford F (1938) The law of anomalous numbers. Proc Am Phil Soc 78: 551–572.

4. Stigler SM (1980) Stigler's law of eponymy. T New York Acad Sci 39: 147–157. doi: 10.1111/j.2164-0947.1980.tb02775.x

5. Pinkham RS (1961) On the distribution of first significant digits. Ann Math Statist 32: 1223–1230. doi: 10.1214/aoms/1177704862

6. Hill TP (1995) A statistical derivation of the significant-digit law. Stat Sci 10: 354–363.

7. Raimi R (1969) On the distribution of first significant figures. Am Math Mon 76: 342–348. doi: 10.2307/2316424

8. Nigrini MJ (2012) Benford's Law: Applications for forensic accounting, auditing, and fraud detection. Hoboken: John Wiley & Sons.

9. Kossovsky AE (2014) Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications. Singapore: World Scientific.

10. Berger A, Hill TP (2011) A basic theory of benford's law. Probab Surv 8: 1–126. doi: 10.1214/11-PS175

11. Lemons DS (1986) On the numbers of things and the distribution of first digits. Am J Phys 54: 816–817. doi: 10.1119/1.14453

12. Wlodarski J (1970) Fibonacci and lucas numbers tend to obey benford's law. Fibonacci Quarterly 9: 87–88.

13. For an excellent list of articles on Benford's Law see the Benford Online Bibliography: www.benfordonline.net, accessed 2015 January 7.

14. Nigrini MJ, Miller SJ (2007) Benford's law applied to hydrology data—results and relevance to other geophysical data. Math Geol 39: 469–490. doi: 10.1007/s11004-007-9109-5

15. Buck B, Merchant AC, Perez SM (1993) An illustration of benford's first digit law using alpha decay half lives. Eur J Phys 14: 59–63. doi: 10.1088/0143-0807/14/2/003

16. Shao L, Ma BQ (2009) First digit distribution of hadron full width. Mod Phys Lett A 24: 3275–3282. doi: 10.1142/S0217732309031223

17. Shao L, Ma BQ (2010) The significant digit law in statistical physics. Physica A 389: 3109–3116. doi: 10.1016/j.physa.2010.04.021

18. Formann AK (2010) The newcomb-benford law in its relation to some common distributions. PLoS ONE 5: e10541. doi: 10.1371/journal.pone.0010541 PMID: 20479878

19. Giles DE (2007) Benford's law and naturally occurring prices in certain ebay auctions. Appl Econ Lett 14: 157–161. doi: 10.1080/13504850500425667

20. Busta B, Sundheim R (1992) Tax return numbers tend to obey benford's law. Center for Business Research Working Paper No W93-106-94, St Cloud State University, Minnesota.

21. Hickman MJ, Rice SK (2010) Digital analysis of crime statistics: Does crime conform to benford's law? J Quant Criminol 26: 333–349. doi: 10.1007/s10940-010-9094-6

22. Durtschi C, Hillison W, Pacini C (2004) The effective use of benford's law to assist in detecting fraud in accounting data. J Forensic Account: 17–34.

23. Diekmann A, Jann B (2010) Benford's Law and Fraud Detection: Facts and Legends. German Economic Review 11: 397–401. doi: 10.1111/j.1468-0475.2010.00510.x

24. Raimi RA (1985) The first digit phenomenon again. Proc Amer Philosophical Soc 129: 211–219.

25. Ashcraft MH, Christy KS (1995) The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. J Res Math Ed 26: pp. 396–421. doi: 10.2307/749430

26. Hamann MS, Ashcraft MH (1986) Textbook presentations of the basic addition facts. Cognition and Instruction 3: 173–202. doi: 10.1207/s1532690xci0303_2

27. There are many instructional examples to be found online. See, for example, the following web-sites accessed 2015 January 7: Psych Web, www.psywww.com/selfquiz/aboutq.htm; University of Central Floridas Student Academic Resource Center, http://sarconline.sdes.ucf.edu/?p=658; University of Lethbridges Faculty of Education Student Handout, http://www.uleth.ca/edu/runte/tests/take/mc/how.html#Tricks

28. Attali Y, Bar-Hillel M (2003) Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. J Educ Meas 40: 109–128. doi: 10.1111/j.1745-3984.2003.tb01099.x

29. Alexander JC (2009) Remarks on the use of benford's law. Working paper, Case Western Reserve University, Department of Mathematics and Cognitive Science.

30. Indeed, in our study we find the χ^2 test to suffer from such excess power. Generally we find that our small data sets such as Knight and Skoog end-of-chapter answers pass the χ^2 threshold test (15.5 for $p<0.05$ and 8 degrees of freedom), while the larger data sets such as the testbank distractors and Young end-of-chapter answers marginally fail the test despite the fact that they visually appear to conform to Benford's Law as well or better than the former.

31. Hoppe FM (2013) Benford's law and distractors in multiple choice exams. arXiv:13117606.

32. Rodriguez MC (2005) Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educ Meas Iss Prac 24: 3–13. doi: 10.1111/j.1745-3992.2005.00006.x

33. Adhikari AK, Sarkar BP (1968) Distribution of most significant digit in certain functions whose arguments are random variables. Sankhya B 30: 47–58.

34. Barlow JL, Bareiss EH (1985) On roundoff error distributions in floating point and logarithmic arithmetic. Computing 34: 325–347. doi: 10.1007/BF02251834