Leftmost Digits' Effect on Numerical Cognition in College Basketball Recaps

Joshua S. Saldaña

Honors Thesis

Rutgers-Camden School of Business

May 2009

Professor Robert Schindler, Thesis Advisor

Professor Julie Ruth, Second Reader

Leftmost Digits' Effect on Numerical Cognition in College Basketball Recaps

Joshua S. Saldaña

Honors Thesis

Rutgers-Camden School of Business

May 2009

Robert Schindler

Professor Robert Schindler, Thesis Advisor

Julie G Ruth

Professor Julie Ruth, Second Reader

Table of Contents

Abstract - 1 -

Previous Research - 3 -

Overview and Hypothesis - 6 -

Methods - 7 -

Selection Procedure - 9 -

Results - 10 -

Descriptive Results - 10 -

Testing Leftmost Digit Effects - 11 -

Unified Measure of Leftmost Digit Difference - 12 -

Measurement of Descriptor Strength and Tests - 12 -

Discussion - 18 -

What We Have Learned - 19 -

Further Research - 20 -

Conclusions - 21 -

References - 22 -

Appendix - 23 -

Exhibit A - 24 -

Exhibit B - 25 -

Exhibit C - 26 -

Exhibit D - 27 -

Exhibit E - 28 –

Abstract

The "leftmost digit effect," the perception that two prices with differing leftmost digits are farther apart than they really are, has been the subject of many studies over the past decade. This study examines the effect of leftmost digits on the perception of numbers in news reporters' descriptions of college basketball games. It was hypothesized that when the leftmost digit difference between the two scores is high, the descriptors (words/short phrases used by reporters to describe the magnitude of the score difference) used would be significantly more powerful than when the leftmost digit difference is low. The sample included 1,491 basketball game recaps, containing over 150 unique descriptors. The top eleven most frequently used descriptors accounted for 88% of all cases. Descriptors were coded for strength both subjectively and objectively. Results indicated that, when descriptor strength is coded subjectively on a strength scale of one to six, the pattern of descriptor incidence was in the predicted direction only half the time. When coded objectively (as "strong" or "weak"), 47% of the top eleven descriptors occurred in the predicted direction. When tested as a group, "strong" and "weak" descriptors both occurred in the predicted direction, but the differences were not statistically significant ($p = 0.16$). Overall, the results were inconclusive and further research is needed.

Leftmost Digits' Effect on Numerical Cognition in College Basketball Recaps

The way consumers perceive prices is of great interest to marketers and managers alike. One observation that has been the subject of many studies is that prices ending with the numbers 0, 5, and 9 occur much more frequently than prices ending in any other single digit (Kreul 1982; Rudolph 1954; Twedt 1965). One study in particular analyzed the rightmost digits of over 1400 retail advertisements to find that over 30% of the displayed prices ended in the number 9 (Schindler & Kirby 1997).

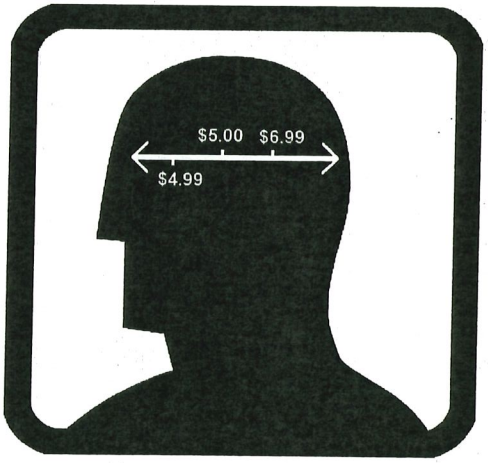
The high occurrence of 9 as the rightmost digit, or ending digit, in pricing has been attributed to many factors, including consumers' excessive attention to a price's leftmost digit. In the above studies, the "leftmost digit" refers to the first single digit in the price, such as the four in $4.99, the one in $19.99, and the two in $299.99. For such prices, the leftmost digit is lowered because of the use of the 99-ending. If the price were one cent higher, the leftmost digit would be one number higher (e.g., $4.99 + .01 = $5.00). Lowering the leftmost digit of a price by lowering the actual price by one cent is widely used throughout the world, but it raises the question, who is really fooled by this pricing tactic?

The simple answer to this question is that people often behave as if they are fooled. A catalog experiment by Schindler and Kibarian (1996) gives some evidence of this. In the study, the actual buying behavior of 60,000 people was studied in relation to the prices they were exposed to. It was found that purchases increased by 8% when the consumers received a catalog containing prices all ending in 99, compared to the identical catalog containing all prices one cent higher than those in the 99-ending price catalog. This again, raises the question: why were so many people affected by this tiny, one-cent price difference?

Previous Research

There have been several studies directly testing whether a difference in the leftmost digits of a set of numbers affects the way they are perceived. One thought as to why this apparent phenomenon exists is the idea of numerical cognition. Simply put, numerical cognition is the idea that the formal arithmetic properties of numbers are different from the properties that we give numbers in our minds. For example, saving $2 on a box of cereal feels a lot better to a consumer than saving $2 on a refrigerator. The *same* amount of money ($2) is perceived to have two different values based on the situation in which it is used. In a similar way, then, a 99-ending and 00-ending price may be considered very different, even though the actual numerical difference between them is small.

Thomas and Morwitz (2005) suggested that consumers can respond to very small price changes, even as small as one cent. They studied the analog model (or holistic model) of numerical cognition, which suggests that when a person encounters two numbers to be compared, he or she perceives the value of the numbers by placing them on an internal analog magnitude scale, much like the picture to the right (Thomas & Morwitz 2005). They used this model to test the effect that nine endings have on price perception.



It was reasoned that the left-to-right processing of numerical symbols can affect the internal "mapping" process and thus distort the analog price magnitude. This is because a number's magnitude is created the instant the first (leftmost) digit is encountered (Dehaene 1997). Because a nine ending (verses a zero ending) changes the leftmost digit, it should affect the conversion of numbers to internal magnitudes.

One of their hypotheses stated that the left-digit effect and other effects in their study are not restricted to the encoding of prices, but they should also affect other multi-digit numbers if it is due, at least in part, to left-to-right processing. In other words, the left-digit effect should be present in other situations where comparing numbers is necessary.

In one of Thomas and Morwitz' experiments, each participant saw two pairs of pens with their respective prices, and the price endings were manipulated to be either a 99-ending or a 00-ending. Participants were asked whether they thought the price of the target pen was "high." It was determined that, "nine endings increased the difference in perceived price magnitude ... only when the dollar digit changed," supporting the existence of the left-digit effect (Thomas and Morwitz 2005).

Manning and Sprott (2009) expanded on Thomas and Morwitz' research by testing whether the left-digit effect influences choice in shopping scenarios. It was proposed that if all other factors are held constant, the greater the perceived price difference, the greater the likelihood that consumers will choose the lower priced option.

Manning and Sprott's participants were shown two comparable pens, paired with a brief description and one of the following three sets of prices: \$2/\$2.99, \$1.99/\$3, or \$2/\$3. To justify the higher price in the participant's mind, in each case the higher priced pen was described in a more refined way, but the two pens were essentially the same. Participants were asked to indicate which pen they would buy. When the leftmost digit remained the same (\$2 / \$2.99), a lower portion of participants chose the lower-priced pen than for the other price sets. The lower-priced pen was chosen the most frequently when the difference between the leftmost digits was the greatest (\$1.99 / \$3).

Bizer and Schindler (2005) indirectly tested the left-digit effect by testing whether consumers tend to drop off prices' rightmost digits. The "drop-off mechanism" is the idea that for an item

priced at $3.99, people may treat the ending digits as zeros, or considerably less then 99, and exhibit behavior consistent with their thought that the item costs about $3, or somewhere between $3 and $4.

Participants were shown an item's price and asked to estimate, without calculating mathematically, how many of these items could be bought for $73. The price's ending digits were manipulated and participants were shown either a price with a 99-ending or with the one-cent higher 00-ending. If the drop-off mechanism (and thus the left-digit effect) appeared in this test, then the participants would, "give higher estimates of the number of items that could be purchased for $73 when the items' price has a 99-ending than when the item's price has a 00-ending" (Bizer & Schindler 2005). Responses that indicated the rightmost digits were dropped off were coded as committing the "drop-off error." For example, for an item priced at $2.99, the correct number of items that could be purchased for $73 is 24.4. If the ending digits were dropped off, making the price now $2, the respondent's estimate would be around 37. In this case, responses between 26 and 37 would be considered as committing the "drop-off error."

Bizer and Schindler's hypothesis was supported; respondents provided higher estimates for items with a price ending in 99. The mean item estimates (across all price levels) for 99 ending prices was 18.33 as opposed to 17.5 for 00 ending prices, a statistically significant difference of 4.5%.

In his pricing textbook, Monroe (1979) noted that the leftmost digit of a price is one of the main cues that affects consumers' price perception. Monroe, also a college basketball fan, reportedly observed that even when two basketball scores were exactly the same distance apart, news reporters used stronger language to describe games where the leftmost digit difference between the two scores was higher. This led him to create the following exercise:

Example 1. Listed below are two sets of prices. Without calculating the differences, which pair of prices, set A or set B *appears* to be farther apart?

A		B	
$83	$67	$87	$71

Example 2. Listed below are three sets of prices. Without performing any calculations, which set of prices *appears* to have the highest average?

A	$87	$81	$75	$69
B	$89	$79	$73	$71
C	$84	$81	$75	$72

Without calculating the answers, one would agree that the numbers in Set A in Example 1 appear to be farther apart. In actuality, however, the numbers in Set A and Set B are exactly the same distance apart. The numbers in Set A may appear to be farther apart because the leftmost digits in Set A are two digits apart whereas Set B's leftmost digits are only one digit apart.

The goal of this study is to provide some further evidence for this apparent phenomenon. Monroe got the idea for making these two-number comparisons when hearing reports of basketball scores. This study will examine whether the leftmost digits of basketball scores have any effect on the way they are described by reporters.

Newspaper reporters who write recaps for basketball games may see and compare hundreds of sets of numbers per month (each basketball score containing two multi-digit numbers, e.g., 65-72), so if this behavioral phenomenon exists, it should be present in the reporters' written descriptions of the basketball games. The language used by the reporter in the game recap will be analyzed to see if a game with a certain point spread is described differently when the leftmost digit difference is large than when it is small.

Overview and Hypothesis

Given Monroe's intuitions and the supporting evidence, one would expect that the left-digit

effect will be present in news reporters' descriptions of basketball games. Game scores with larger leftmost digit differences make the difference between the two scores in the recap seem larger, and thus the win or loss looks more dramatic. Scores with a larger leftmost digit difference should thus be described with stronger language, more appropriate for a larger difference, than games of the same point difference where the leftmost digits are the same. Formally:

> *Hypothesis:* For a given point spread (actual difference), a basketball score pair with a higher leftmost digit difference will be described using more powerful language than a score pair with a lower leftmost digit difference.

Methods

In order to test our hypothesis, we needed to determine the actual point differences, or point spreads, that would be selected for the study. Typical college basketball score spreads range from one to nineteen points, but scores with a point difference of 4, 5, 6, 14, 15, 16 were specifically chosen for this investigation (see Appendix, Exhibit A). These score differences were chosen because they have roughly the same number of possible ending digit pairs for larger leftmost digit differences (LDDs) as for smaller ones. Consider the following table:

TABLE 1

Possible Score Combinations For a Given Decade (60s)					
4-Point Spread		**5-Point Spread**		**6-Point Spread**	
LDD = 0	**LDD = 1**	**LDD = 0**	**LDD = 1**	**LDD = 0**	**LDD = 1**
60-64	66-70	60-65	65-70	60-66	64-70
61-65	67-71	61-66	66-71	61-67	65-71
62-66	68-72	62-67	67-72	62-68	66-72
63-67	69-73	63-68	68-73	63-69	67-73
64-68		64-69	69-74		68-74
65-69					69-75

The table above shows three examples of actual point differences within a given decade (60s). For a basketball score with a 5-point difference, or 5-point spread, the leftmost digit difference can be either zero or one (e.g. LDD = 0: 60-65 or LDD = 1: 65-70). The probability is exactly the same that a case where the LDD = 1 will be found compared to the probability that a case where LDD = 0 will be found. Point spreads four and six have similar probabilities of LDD = 0 and LDD = 1, so they were considered acceptable for collection in this study.

Games with point spreads above 20 were avoided because these are relatively rare and may not exist in enough cases to establish a solid conclusion. Also, games with point spreads of 10 and 20 were avoided because games with these point spreads will always have the same leftmost digit difference (1 and 2 respectively) and thus will introduce a confounding factor. The point spreads 1-3, 7-9, 11-13, and 17-19 were avoided because they also introduce a confounding. Contrary to the point spreads shown in Table 1, the number of possible score combinations is greatly unbalanced in cases where LDD = 1 versus when LDD = 0. These scores have unequal probabilities that cases with different LDDs will be found. Again, using a given decade (60s), Table 2 illustrates two examples of confounding:

TABLE 2

Examples of Confounding Within A Given Decade (60s)			
1-Point Spread		**9-Point Spread**	
LDD = 0	**LDD = 1**	**LDD = 0**	**LDD = 1**
60-61	69-70	60-69	61-70
61-62			62-71
62-63			63-72
63-64			64-73
64-65			65-74
65-66			66-75
66-67			67-76
67-68			68-77
68-69			69-78

The six chosen score differences (4-6, 14-16), each with two conditions (left-digit change small and left-digit change large), makes 12 separate conditions that must be filled. Consistent with past research on leftmost digit effects on advertised prices (Schindler 2009), a sample size of 1500 was deemed adequate to test our hypothesis.

Selection Procedure

Because of the nature of this study, NCAA basketball games (National Collegiate Athletic Association) were studied. Collegiate games typically are not as high-scoring as NBA games, usually not going into triple digits, so they fit the requirements of this study.

The *New York Times* online archives were utilized to collect the sample data. A search was run for every article with the headline "College Basketball," which is the standard *New York Times* headline for collegiate basketball reports. In the data collection, both men's and women's teams were included, and pre- and post-season games were included in with the regular season games.

The search listed the articles by date of publication, starting from the most recent news article. All college basketball scores within our predetermined parameters were collected, with no geographical constraints. When a game was found that had an acceptable point spread (i.e. 4-6, 14-16), it was copied to a word processing file. For each point spread, the selection procedure alternated so that every other collected game had the same leftmost digit difference. For instance, if the first game collected was 80-84, the next game collected with a point spread of 4 had to have a leftmost digit change, such as 88-92, 69-73, etc., and the process continued to the next 4-point-spread game collected, which would not have a leftmost digit change. While waiting for an acceptable case of 4-point spread, data collection was simultaneously being done

for other acceptable point spreads in the same manner. This procedure insured that an equal number of large and small leftmost digit differences would be collected for each point spread.

Data collection alternated in this way until almost 1,500 cases were collected. There are over 300 collegiate teams in the NCAA, and each team plays about 30 games per season, plus there were some post-season games, so the *New York Time*'s NCAA reports provided adequate data for the study (see Appendix, Exhibit B, for examples of selected recaps).

Results

After 1,491 cases were collected, the data were analyzed using the *Statistical Package for the Social Sciences* (SPSS). For each set of point spreads ("Low point spreads" being 4-6, and "High point spreads" being 14-16), the number of cases of low versus high leftmost digit differences were about equal (see Table 3 below).

TABLE 3

Total cases: 1491		
	Low point spreads (4-6)	**High point spreads (14-16)**
Low Leftmost Digit Difference	425	317
High Leftmost Digit Difference	427	322

Descriptive Results

Within the 1,491 cases, there were over 150 unique descriptors, which were single words or short phrases (e.g. "carried past") used to describe the spread of the two basketball scores. Sometimes reports used different forms of the same word. This required that the 150+ descriptors be boiled down to each word/phrase in their most commonly used form, usually the

past-tense, singular form of the word. For example, descriptors like "beat," "beaten," and "beating" were all categorized under the word "beat," and "rally," "rallied," "rallied past," and "rallied over" were all categorized as "rallied."

This coding process resulted in exactly 80 different descriptors. These 80 are used in all further analyses. Shown below is a table of the top eleven most commonly used descriptors. The top eleven were chosen for the split because there is a large break in the number of uses for the rest of the descriptors (see Appendix, Exhibit C-D). Note that three descriptors make up over half of the sample data, and 14% of the descriptors (11 of 80) make up 88% of the sample data.

TABLE 4

Frequencies of Descriptors

Rank	Descriptor	Count	Percentage	Rank	Descriptor	Count	Percentage
1	Victory	415	27.8	8	Upset	64	4.3
2	Beat	239	16	9	Lift	18	1.2
3	Defeat	138	9.3	10	Snapped Streak	14	0.9
4	Win	133	8.9	11	Rallied	14	0.9
5	Led	132	8.9	*12*	*Other*	*202*	*12.2*
6	Held	77	5.2				
7	Loss	65	4.4		**Total**	**1491**	**100 %**

As can be seen in the table, there were a few descriptors used very frequently, and many others that were used very rarely.

Testing Leftmost Digit Effects

In order to test the hypothesis, the two variables used in this study, the descriptors and the leftmost digit differences, must be converted into measurable scales. To accomplish this, the leftmost digit differences were "unified," so that the low and high point spreads could be comparable. The strength of the 80 descriptors were rated in two ways, one using a subjective rating system, and the other using an objective system.

Unified Measure of Leftmost Digit Difference

In order to test descriptor strength trends according to varying leftmost digit differences, there must be a unification of the two sets of point spreads found within our data sample. For low point spreads (4, 5, and 6), the possible leftmost digit differences are 0 and 1 (see Table 1, page 6). For high point spreads (14, 15, and 16), the possible leftmost digit differences are 1 and 2. To unify the leftmost digit differences for testing the hypothesis, the lowest leftmost digit difference within both sets of point spreads were combined into one value of a new variable and the same was done for the higher leftmost digit differences in the point spread sets. The new variable was called "Unified LM Difference." Leftmost differences 0 from point spreads 4-6 and 1 from point spreads 14-16 were both coded as a "1." Leftmost differences of 1 from point spreads 4-6 and 2 from point spreads 14-16 were both coded as a "2" (see Appendix Exhibit E).

Measurement of Descriptor Strength and Tests

A subjective rating system was created by the researchers to assign a strength level to each of the 80 descriptors. The scale ranged from one to six, six being the strongest strength level. The descriptors assigned to each point in the strength rating scale means that the descriptor fulfills one of six conditions, outlined in Table 5. Two researchers rated each descriptor independently and when ratings differed, they discussed the differences to arrive at a single conclusion.

TABLE 5

Subjective Descriptor Rating Scale		
6	-	Strong descriptor indicating high point spread
5	-	Weak descriptor indicating high point spread
4	-	Strong descriptor indicating moderate point spread
3	-	Weak descriptor indicating moderate point spread
2	-	Strong descriptor indicating low point spread
1	-	Weak descriptor indicating low point spread

After all descriptors were assigned a strength level, the data set was subjected to a cross-tabulation to see if descriptors with higher strength levels are used to describe scores with a higher leftmost digit difference. Results were that only three out of the six strength points were in the predicted direction, and the differences were not statistically significant (see Tables 6 and 7 below).

TABLE 6

Subjective Ratings * Unified LM Difference Crosstabulation

			Unified LM Difference		Total
			1	2	
Subjective Ratings	1	Count	62	64	126
		% within Unified LM Difference	8.4%	8.5%	8.5%
	2	Count	34	32	66
		% within Unified LM Difference	4.6%	4.3%	4.4%
	3	Count	114	128	242
		% within Unified LM Difference	15.4%	17.1%	16.2%
	4	Count	510	502	1012
		% within Unified LM Difference	68.9%	66.9%	67.9%
	5	Count	8	11	19
		% within Unified LM Difference	1.1%	1.5%	1.3%
	6	Count	12	13	25
		% within Unified LM Difference	1.6%	1.7%	1.7%
	Total	Count	740	750	1490
		% within Unified LM Difference	100.0%	100.0%	100.0%

TABLE 7

Chi Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.412[a]	5	0.923
Likelihood Ratio	1.415	5	0.923
Linear-by-Linear Association	0.028	1	0.866
N of Valid Cases	1490		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.44.

After conducting this test, the accuracy of the rating system came into question and the researchers realized the limits of subjective ratings. The next step was to examine the descriptors more objectively.

The descriptors were tested objectively by observing which descriptors were more commonly used in cases with a point spread of 4, 5, or 6 versus cases with a point spread of 14, 15, or 16. This seemed to be the most logical way to begin testing because common sense would tell us that stronger descriptors would occur more frequently to describe games with a higher point spread (see Table 8 below).

TABLE 8

Verbcat * 4-6 vs. 14-16 Crosstabulation

			4-6 vs. 14-16			
			1	2	Total	Difference
Verbcat	Victory	Count	238	177	415	
		% within 4-6 vs. 14-16	**28.0%**	27.7%	27.8%	(0.003)
	Beat	Count	125	114	239	
		% within 4-6 vs. 14-16	14.7%	**17.8%**	16.0%	0.031
	Defeat	Count	77	61	138	
		% within 4-6 vs. 14-16	9.0%	**9.5%**	9.3%	0.005
	Win	Count	71	62	133	
		% within 4-6 vs. 14-16	8.3%	**9.7%**	8.9%	0.013
	Led	Count	70	62	132	
		% within 4-6 vs. 14-16	8.2%	**9.7%**	8.9%	0.015
	Held	Count	63	14	77	
		% within 4-6 vs. 14-16	**7.4%**	2.2%	5.2%	(0.052)
	Loss	Count	34	31	65	
		% within 4-6 vs. 14-16	4.0%	**4.8%**	4.4%	0.008
	Upset	Count	52	12	64	
		% within 4-6 vs. 14-16	**6.1%**	1.9%	4.3%	(0.042)
	Lift	Count	15	3	18	
		% within 4-6 vs. 14-16	**1.8%**	.5%	1.2%	(0.013)
	Snapped Streak	Count	7	7	14	
		% within 4-6 vs. 14-16	.8%	**1.1%**	.9%	0.003
	Rallied	Count	8	6	14	
		% within 4-6 vs. 14-16	.9%	.9%	.9%	(0.000)
	Z-other	Count	91	91	182	
		% within 4-6 vs. 14-16	10.7%	**14.2%**	12.2%	0.035
Total		Count	851	640	1491	
		% within 4-6 vs. 14-16	100.0%	100.0%	100.0%	

The top eleven most frequently occurring descriptors, and the other 69 descriptors (labeled Z-other) are listed in ascending order of frequency. If a higher percentage (circled) of a descriptor

is observed in cases with a higher point spread (14, 15, and 16), then the descriptor is considered "strong." Likewise, if the higher percentage of a descriptor is seen in cases with low point spreads (4, 5, and 6), then the descriptor is considered "weak." For example, the descriptor "Victory" occurs more frequently (28%) in cases with a point spread of 4, 5, and 6, so for subsequent tests, it was considered "weak."

The 69 least frequently occurring group of descriptors contains both "strong" and "weak" individual descriptors, but this category was tested as a whole. Descriptors in this group are present more (14.2%) in cases with a point spread of 14, 15, and 16, and were thus considered "strong." Within this category, 37 of the descriptors occurred more than once (but fewer than 14 times), and the other 32 occurred only once. Table 9 below shows the top ten "strong" and "weak" low-occurring descriptors.

TABLE 9

Low-Occurring Descriptors	
Top 10 Strongest	**Top 10 Weakest**
Clubbed	Avoid Losing
Whipped	Clinch
Smacked	Comeback
Thwarted	Deflated
Rolled	Dismissed
Snuffed	Dropped
Disposed of	Edged
Overpowered	Eliminated
Dispatched	Escape
Routed	Fall

There were also some descriptors that seemed to focus more on the surprise of the final score rather than the actual magnitude of the score difference. Descriptors like "Surprised," "Overtook," and "Comeback" all raise this issue of the surprise factor. Outside of the low-occurring descriptors, "Upset" is in the top eleven most frequently occurring descriptors, and it

also implies a measure of surprise. In basketball, unexpected winners, or underdogs, generally win by a low number of points, and most of the time it occurs as a last minute scoring drive. Reasonably, this is why it occurs more in cases with a small point spread and is considered "weak."

Table 10 below shows the crosstabulation of the top eleven descriptors by the unified leftmost differences.

TABLE 10

Verbcat * Unified LM Difference Crosstabulation

			Unified LM Difference		Total	Strong/Weak	In Predited Direction?
			1	2			
Verbcat	Beat	Count	112	127	239	Strong	Yes
		% within Unified LM Difference	15.1%	**16.9%**	16.0%		
	Defeat	Count	68	70	138	Strong	Yes
		% within Unified LM Difference	9.2%	**9.3%**	9.3%		
	Led	Count	69	63	132	Strong	No
		% within Unified LM Difference	**9.3%**	8.4%	8.9%		
	Loss	Count	40	25	65	Strong	No
		% within Unified LM Difference	**5.4%**	3.3%	4.4%		
	Snapped Streak	Count	8	6	14	Strong	No
		% within Unified LM Difference	**1.1%**	.8%	.9%		
	Win	Count	67	66	133	Strong	No
		% within Unified LM Difference	**9.0%**	8.8%	8.9%		
	Z-other	Count	79	103	182	Strong	Yes
		% within Unified LM Difference	10.7%	**13.7%**	12.2%		
	Held	Count	45	32	77	Weak	Yes
		% within Unified LM Difference	**6.1%**	4.3%	5.2%		
	Lift	Count	6	12	18	Weak	No
		% within Unified LM Difference	.8%	**1.6%**	1.2%		
	Upset	Count	28	36	64	Weak	No
		% within Unified LM Difference	3.8%	**4.8%**	4.3%		
	Victory	Count	212	203	415	Weak	Yes
		% within Unified LM Difference	**28.6%**	27.1%	27.8%		
	Rallied	Count	7	7	14	-	Equal
		% within Unified LM Difference	.9%	.9%	.9%		
Total		Count	741	750	1491		
		% within Unified LM Difference	100.0%	100.0%	100.0%		

The descriptors are given their objectively assigned strength in the "Strong/Weak" column, and based on that, it can be easily determined whether the descriptor goes in the predicted direction. "Strong" descriptors are predicted to occur more frequently when the Unified LM Difference is 2, and "weak" descriptors are predicted to occur more frequently when the Unified

LM Difference is 1. As can be seen in Table 10, there are five descriptors that occur in the predicted direction, and six that do not. Note that the highest percent difference between the leftmost differences that occurred in the predicted direction was the Z-other category with a 3.0% difference.

Again, the failure of the descriptor "Upset" to be in the predicted direction raises the issue of surprise versus magnitude of the point spread. "Upset" occurred more frequently in low point spreads, and was coded as "weak," giving evidence that the surprise factor outweighs the magnitude of the point spread. However, here "Upset" occurs more frequently in cases of higher leftmost differences, so it leaves open the question of how the element of surprise should be taken into account.

Lastly, the objectively assigned descriptor strengths were subjected to a crosstabulation again involving the Unified LM Difference. In the previous test, the descriptors were looked at individually as to how frequently they occur per (unified) leftmost digit, but in this test, shown in Table 11, it is overall descriptor strength and Unified LM Difference that are crosstabulated.

TABLE 11

Descriptor Strength * Unified LM Difference Crosstabulation

			Unified LM Difference		Total
			1	2	
Descriptor Strength	Weak	Count	263	247	510
		% within Descriptor Strength	**51.6%**	48.4%	100.0%
		% within Unified LM Difference	**35.5%**	32.9%	34.2%
	Strong	Count	478	503	981
		% within Descriptor Strength	48.7%	**51.3%**	100.0%
		% within Unified LM Difference	64.5%	**67.1%**	65.8%
	Total	Count	741	750	1491
		% within Descriptor Strength	49.7%	50.3%	100.0%
		% within Unified LM Difference	100.0%	100.0%	100.0%

The results of this test showed that the descriptors, being objectively assigned their strengths, both go in the predicted direction. "Weak" descriptors occurred more frequently when the leftmost difference was low, as shown by their row and column percents, 51.6% and 35.5%. Likewise the "strong" descriptors occurred more frequently in cases of higher leftmost digit differences, with a 2.6% difference within the level of descriptor strength.

This test was also the closest, of any of the tests conducted, to being a statistically significant result. The chi-square tests shown in Table 12 indicates that in a one-sided Fisher's exact test, $p = 0.162$, which is the closest we have come to obtaining a $p = .05$ in any of our results' chi-square tests.

TABLE 12

Chi Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.085[a]	1	0.298		
Continuity Correction[b]	0.974	1	0.324		
Likelihood Ratio	1.085	1	0.298		
Fisher's Exact Test				0.300	0.162
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 253.46.					
b. Computed only for 2x2 table.					

Discussion

Descriptors were coded subjectively and objectively, but none of the resulting differences were found to be statistically significant. When coded subjectively on a strength scale of one to six, three out of six of the strength points were in the predicted direction. When coded objectively, based on the set of point spreads in which the descriptors occurred more frequently, the results still were inconclusive. When tested individually by descriptor, five out of twelve, or 41.7% of descriptors were found to be in the predicted direction. When tested by strengths as a whole, both strength levels, "strong" and "weak" were in the predicted direction. "Strong"

descriptors were used more frequently to describe scores with a high leftmost digit difference, and "weak" descriptors were used more frequently to describe scores with a low leftmost digit difference.

What We Have Learned

Through this study, we have learned that descriptor strength is related to the point spread, i.e., the actual point difference. Stronger words are used in cases with higher actual point differences, and the inverse is equally true. This validates the methods used in this study to both collect and code the descriptors. Also, we noted that the degree of surprise in a game's final scores may be a factor that affects descriptor use. The effect of the degree of surprise versus the actual magnitude of the final scores must be considered in order to test the descriptors accurately.

We have also learned that the "80-20 rule" applies to basketball score descriptors. The 80-20 rule is the basic behavioral observation that 80% of an effect is likely to come from 20% of its causes. This rule is frequently applied in business settings to income; that 80% of a company's income will generally come from the top 20% of the company's clients. This rule also is seen other situations, like the subject of this study. Eighty percent of our data set consisted of about the top 20% of our descriptors.

Besides the "80-20 rule," we note that our data set also follows Zipf's law, which predicts that out of a population of N elements, the frequency (P_n) of occurrence of the n^{th} ranked item is:

$$P_n \sim 1/n^a$$

Zipf's law states that given some body of natural language, the frequency of any word is inversely proportional to its rank in the frequency table (Black 2006). So the most frequently occurring word will occur approximately twice as often as the second most frequent word, which

occurs twice as often as the fourth most frequent word, and so on (see Appendix, Exhibit D). Learning that this law applies to basketball descriptors raises the question, in what other situations can Zipf's law be found?

Further Research

There are additional analyses that could have been conducted. For instance, the 69 low-occurring descriptors were tested only as a group. Even though this group of descriptors accounts for less than 15% of the total sample, it comprises 86% of the total number of descriptors. Low-occurring descriptors could be tested by sampling the top ten strongest descriptors and the top ten weakest descriptors in the group, and then crosstabulating them by the Unified LM Difference. That may add some insight into the trends of low-occurring words, or lead to further testing.

There also may be other factors influencing numerical cognition in this instance. News reporters who specifically report on college basketball see hundreds of games per year, and so their cognition of numbers may be more advanced than the holistic model, which means that they may analyze the spread, actively compare it to other scores they have logged into their mental numerical spectrum, and then think very hard about how they will describe the spread, based on their own past trends. Their minds may also have become so accustomed to seeing a certain set of point spreads, like the ones in our study, that the reporters' minds have become too crowded with logged spreads, so that only ones that differ greatly receive any special descriptor use. This means that a point spread must be past a certain threshold in the cognitive model to be described according to the trends predicted in this study. Therefore, a study of unusually high point spreads could be undertaken, but it may be difficult to obtain enough cases to obtain reliable results.

The one major issue in this investigation was that the study necessitated the collection of large amounts of data, where each case was tediously coded for a number of properties. The amount of time allotted to this study did not do the study justice compared to the sheer amount of data needed and the amount of detail in the coding process. The *New York Times* archives had multiple cases per page, so each page had to be checked for valid cases. If, in the future, there is a way to automate the process of collecting data, or a way to obtain secondary data from either the NCAA or the *New York Times*, it might be possible to do more justice to a test of this study's hypothesis.

Conclusions

Although the results found in this study were inconclusive, the trends were in the predicted direction and some differences were even within reach of statistical significance. However, one should not assume that plausible hypotheses are necessarily correct. With any psychological idea, such as numerical cognition, if it is to be tested against real-world data, one must carefully structure the situation of study and also have a large amount of data. Validly testing an hypothesis against naturally occurring real-world behavior is a steep undertaking, but it can be done, given the right tools and time.

References

Bizer, George Y. and Robert M. Schindler (2005). "Direct Evidence of Ending-Digit Drop-Off in Price Information Processing." *Psychology & Marketing* 22: 771-83.

Black, Paul E. (2006). "Zipf's law." *National Institute of Standards and Technology*. 24 Apr. Accessed May 2009, <http://www.itl.nist.gov/div897/sqg/dads/HTML/zipfslaw.html>.

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

Kreul, Lee M. (1982). Magic Numbers: Psychological Aspects of Menu Pricing. *Cornell Hotel and Restaurant Administration Quarterly*, 23 (August): 70-75.

Rudolph, Harold J. (1954). Pricing for Today's Market. *Printers' Ink*, 247 (May 28): 22-24.

Manning, Kenneth C. and David E. Sprott (2009). "Price Endings, Left-Digit Effects and Choice." *Journal of Consumer Research*, in press.

Monroe, Kent B. (1979). *Pricing: Making Profitable Decisions*. New York: McGraw-Hill.

Schindler, Robert M. and Patrick N. Kirby (1997). "Patterns of Rightmost Digits Used in Advertised Prices: Implications for Nine-Ending Digits." *Journal of Consumer Research* 24: 192-201.

Schindler, Robert M. and Thomas M. Kibarian (1996). "Increased Consumer Sales Response Through Use of 99-Ending Prices", *Journal of Retailing*, 72 (2): 187-99.

Thomas, Manoj, and Vicki Morwitz (2005). "Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition." *Journal of Consumer Research* 32: 54-64.

Twedt, Dik W. (1965). Does the "9 Fixation" in Retail Pricing Really Promote Sales? *Journal of Marketing*, 29 (October): 54 -55.

Appendix

Exhibit A	–	Typical College Basketball Point Spreads
Exhibit B	–	A sample of actual cases taken from 6-point-spread raw data
Exhibit C	–	Descriptor Frequency Table
Exhibit D	–	Descriptor Frequency, plotted
	–	Zipf's Law, plotted
Exhibit E	–	"Unification" of leftmost digit differences

Exhibit A

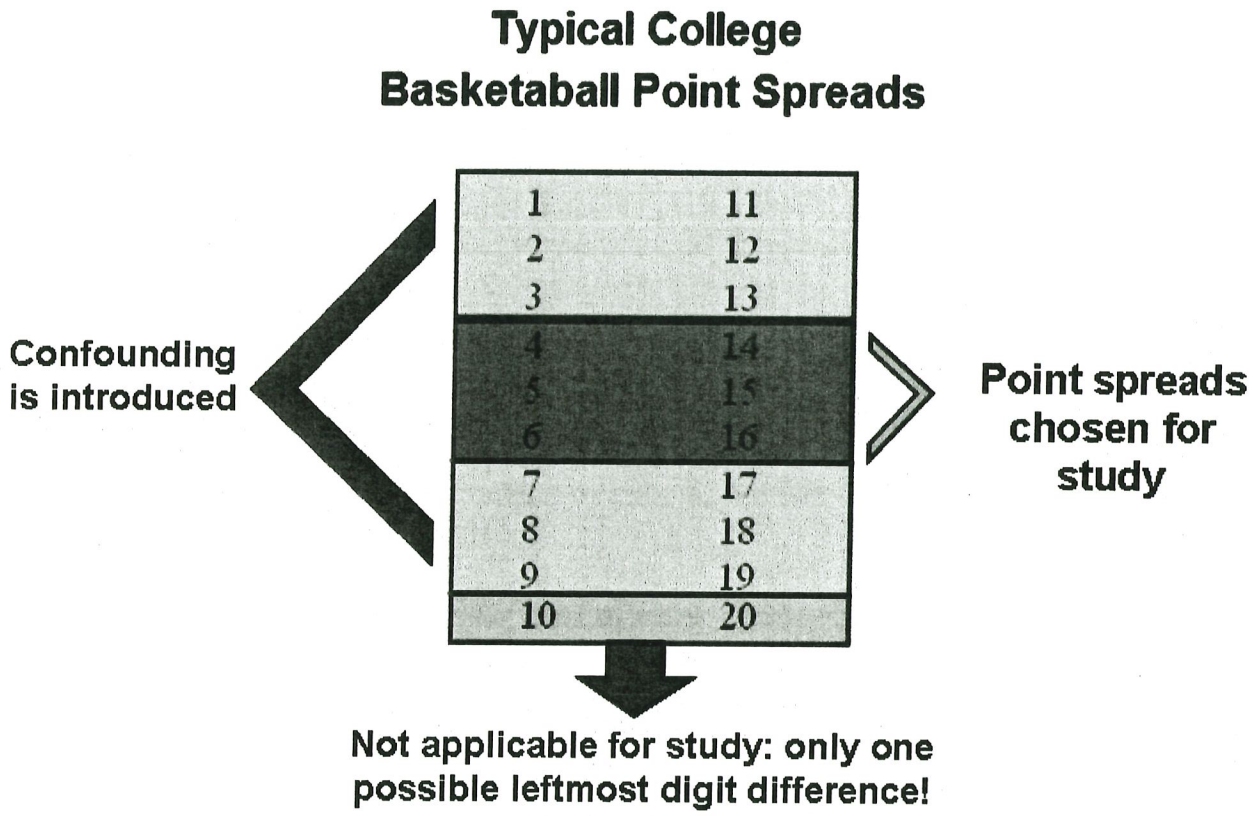

Typical College
Basketaball Point Spreads
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Confounding
is introduced
Point spreads
chosen for
study
Not applicable for study: only one
possible leftmost digit difference!

Exhibit B

Case 16. Alando Tucker scored 15 of his 17 points in the second half and No. 3 Wisconsin held off host Northwestern, 56-50, for its 13th consecutive victory.

Case 23. MICHIGAN 71, MIAMI 65 Daniel Horton scored 19 points and Chris Hunter had 16 as Michigan advanced to the N.I.T. semifinals. Dion Harris and Lester Abram added 15 points apiece to help the Wolverines (21-10) extend the nation's longest postseason winning streak to eight games and improve to 16-0 in home N.I.T. games.

Case 24. Marist defeated ninth-seeded Canisius, 68-62, in the last of the quarterfinals Saturday night. Canisius finished at 9-20. (AP)

Case 264. MARIST 60, ST. PETER'S 54 Kaylen Gregory scored 15 points as Marist beat visiting St. Peter's. Ryan Schneider added 15 points and 6 rebounds for Marist (6-9, 2-1 Metro Atlantic Athletic Conference), and David Devezin had eight assists in his first start since sustaining a heel injury in mid-December. Wesley Jenkins had a game-high 16 points for St. Peter's (4-9, 1-2).

Exhibit C

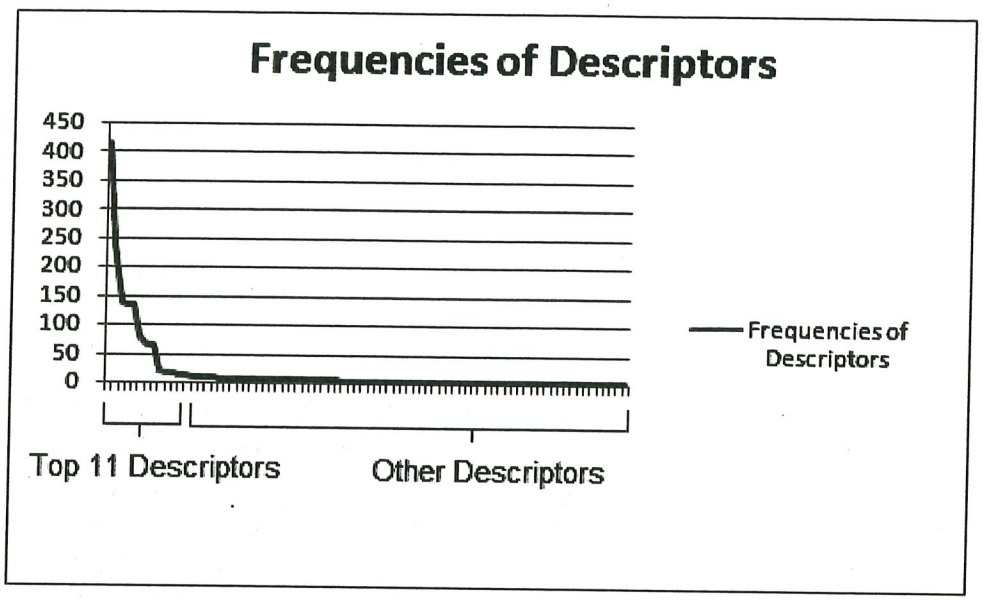
Frequencies of Descriptors

Descriptor	Count	Percentage
Victory	415	27.8
Beat	239	16.0
Defeat	138	9.3
Win	133	8.9
Led	132	8.9
Held	77	5.2
Loss	65	4.4
Upset	64	4.3
Lift	18	1.2
Rallied	14	0.9
Snapped Streak	14	0.9
Survived	11	0.7
Fall	10	0.7
Advance	8	0.5
Ended Streak	8	0.5
Clinch	7	0.5
Overcame	7	0.5
Pulled Away	6	0.4
Rout	6	0.4
Carried	5	0.3
Paced	5	0.3
Put	5	0.3
Took	5	0.3
Downed	4	0.3
Eliminated	4	0.3
Outscored	4	0.3
Rolled	4	0.3
Stunned	4	0.3
Break Open	3	0.2
Dropped	3	0.2
Outlasted	3	0.2
Overpowered	3	0.2
Shocked	3	0.2
Triumph	3	0.2
Turned Back	3	0.2
Whipped	3	0.2
Escape	2	0.1
Extended Streak	2	0.1
Knocked Off	2	0.1
Notched	2	0.1
Prevailed	2	0.1
Remained Undefeated	2	0.1
Shut Down	2	0.1
Stopped	2	0.1
Thwarted	2	0.1
Trounced	2	0.1
Wore Down	2	0.1
Avenged Loss	1	0.1
Avoid Losing	1	0.1
Clubbing	1	0.1
Comeback	1	0.1
Cruised Past	1	0.1
Deflated	1	0.1
Dismissed	1	0.1
Dispached	1	0.1
Disposed of	1	0.1
Eased By	1	0.1
Edged	1	0.1
Overtook	1	0.1
Overwhelmed	1	0.1
Perplexed	1	0.1
Pressed	1	0.1
Pushed	1	0.1
Sink	1	0.1
Slip Past	1	0.1
Smack	1	0.1
Snuffed	1	0.1
Spark	1	0.1
Spoiled	1	0.1
Spur	1	0.1
Spurt	1	0.1
Squeezed By	1	0.1
Stifle	1	0.1
Surge	1	0.1
Surprised	1	0.1
Thrashed	1	0.1
Trimmed	1	0.1
Unable to Hold	1	0.1
Went Ahead	1	0.1
Wilted	1	0.1

Total Descriptors: 80

Total Cases: 1491

Exhibit D


Frequencies of Descriptors
450
400
350
300
250
200
150
100
50
0
Frequencies of Descriptors
Top 11 Descriptors
Other Descriptors

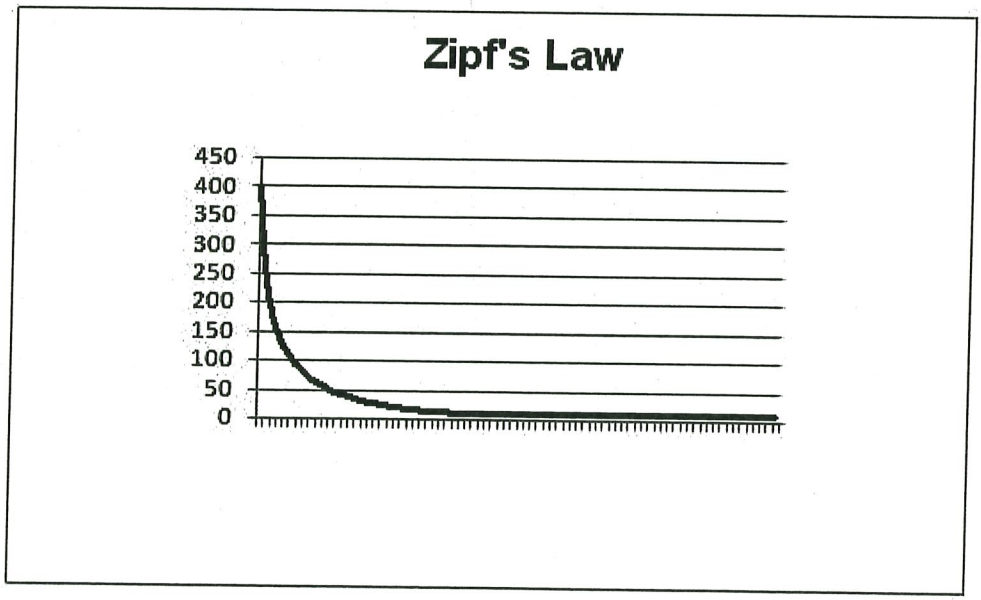
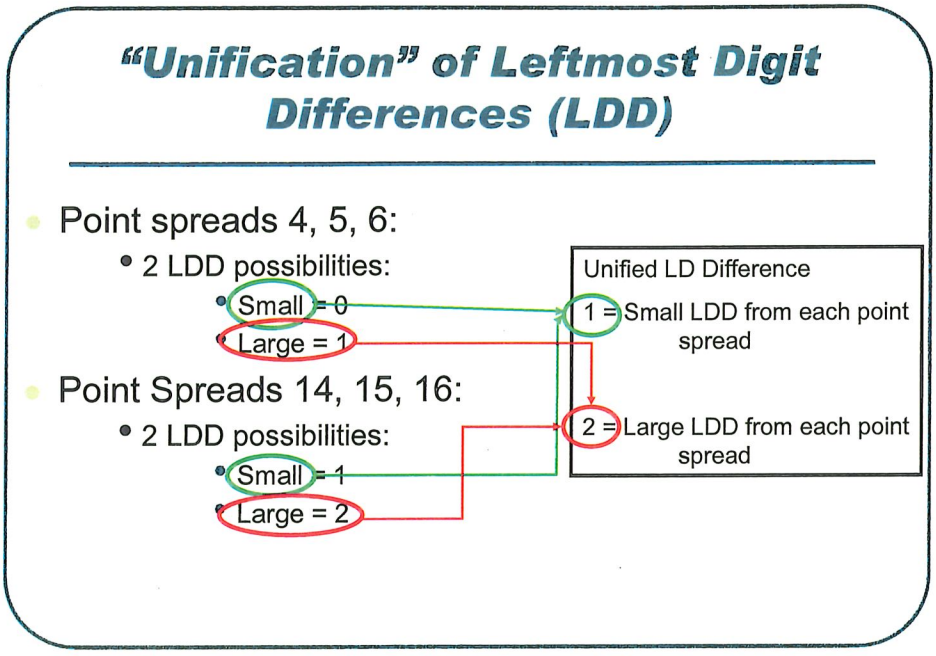

Zipf's Law
450
400
350
300
250
200
150
100
50
0

Exhibit E


"Unification" of Leftmost Digit Differences (LDD)
Point spreads 4, 5, 6:
2 LDD possibilities:
Small = 0
Large = 1
Point Spreads 14, 15, 16:
2 LDD possibilities:
Small = 1
Large = 2
Unified LD Difference
1 = Small LDD from each point spread
2 = Large LDD from each point spread