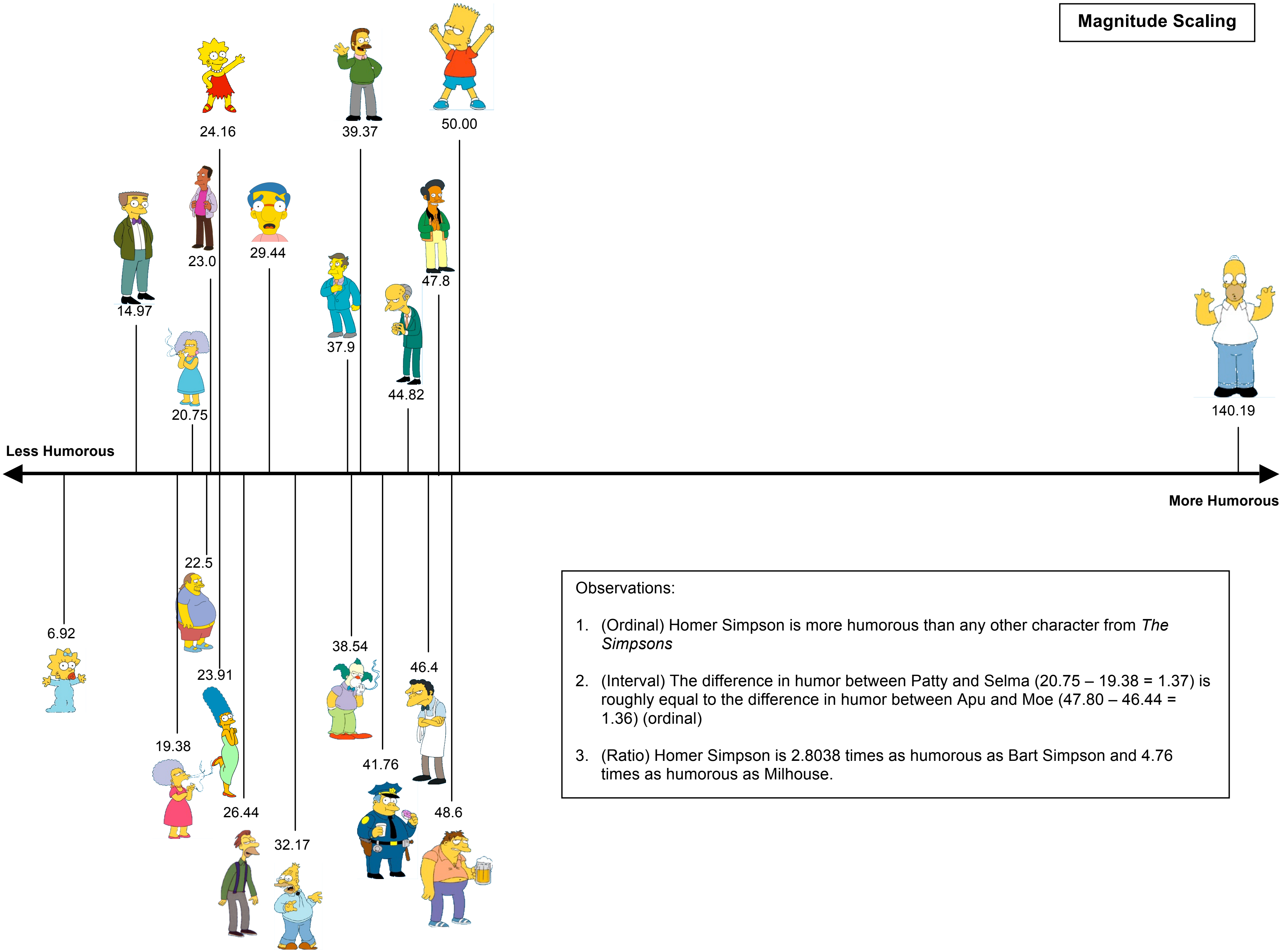
Magnitude Scaling



Observations:

1. (Ordinal) Homer Simpson is more humorous than any other character from *The Simpsons*
2. (Interval) The difference in humor between Patty and Selma (20.75 – 19.38 = 1.37) is roughly equal to the difference in humor between Apu and Moe (47.80 – 46.44 = 1.36) (ordinal)
3. (Ratio) Homer Simpson is 2.8038 times as humorous as Bart Simpson and 4.76 times as humorous as Milhouse.

How hostile is each country towards the United States?

Country	Hostility
Iraq	279.16
Afghanistan	214.96
Saudi Arabia	99.73
Pakistan	99.49
Cuba	94.28
China	89.70
Bosnia	81.41
Russia	65.36
Israel	59.00
Germany	57.45
France	50.00
Japan	48.52
India	44.24
Panama	34.43
Spain	30.48
Mexico	29.13
Canada	18.08
Great Britain	16.35
Sweden	10.31
Australia	6.30

Observations:

1. Iraq is perceived to have 5.6 times as much hostility towards the U.S. as France and 1.3 times as much hostility as Afghanistan
2. As expected, middle-eastern nations were perceived to be the most hostile
3. The ratings for India had the greatest variation, indicating subjects did not agree on India's level of hostility towards the U.S.
4. Spain received unusually high ratings from several respondents (this survey was administered just after the terrorist attack on Spain).
5. English speaking nations had the lowest perceived levels of hostility

How were these results obtained?

83 college students were first administered a simple categorical survey.

```
Instructions:  In this section, you will be presented a list of 20
               countries.  I will ask you to rate each of the
               countries (on a scale from 1 to 7) in terms of the
               amount of hostility you believe they have towards
               the United States.
```

Afghanistan	1	2	3	4	5	6	7
Australia	1	2	3	4	5	6	7
Bosnia	1	2	3	4	5	6	7
Canada	1	2	3	4	5	6	7
…							
Spain	1	2	3	4	5	6	7
Sweden	1	2	3	4	5	6	7

Limitations of Categorical Scales

1. Provide only ordinal level of measurement
2. Category labels influence responses (1 = least hostile; 7 = most hostile)
3. The limited number of categories limits the information we obtain

If these problems are due to the limited number of categories provided to respondents, perhaps we should provide subjects an unlimited number of categories in which to classify stimuli.

In other words, instead of forcing respondents to select a number 1-7, why not allow subjects to assign *any* number to the stimuli? This would seemingly take care of the limitations of categorical scales.

But how would we collect and analyze the data? To answer this question, consider a simple psychophysical experiment…

Experiment: Look at the first line drawn below (the reference line). The length of this line is 50. Your task is to estimate the lengths of the other ten lines on this page by comparing their lengths to the length of the reference line. If you believe a line is exactly twice as long as the reference line, you would estimate the length of that line to be 100. If you believe a line is 5.5 times as long as the reference line, you would estimate the length to be 275 (5.5 x 50 = 275). If you think a line is one-tenth the length of the reference line, you would assign it a value of 5. **Do not actually measure these lines!** Just look at them and write out your estimates.

You may assign any positive number (including fractions or decimals) corresponding to the relative length of the lines. Do not assign negative numbers or the number zero.

Reference Line **Length =** 50

Line #1 **Length =** __________

Line #2 **Length =** __________

Line #3 **Length =** __________

Line #4 **Length =** __________

Line #5 **Length =** __________

Line #6 **Length =** __________

Line #7 **Length =** __________

Line #8 **Length =** __________

Line #9 **Length =** __________

Line #10 **Length =** __________

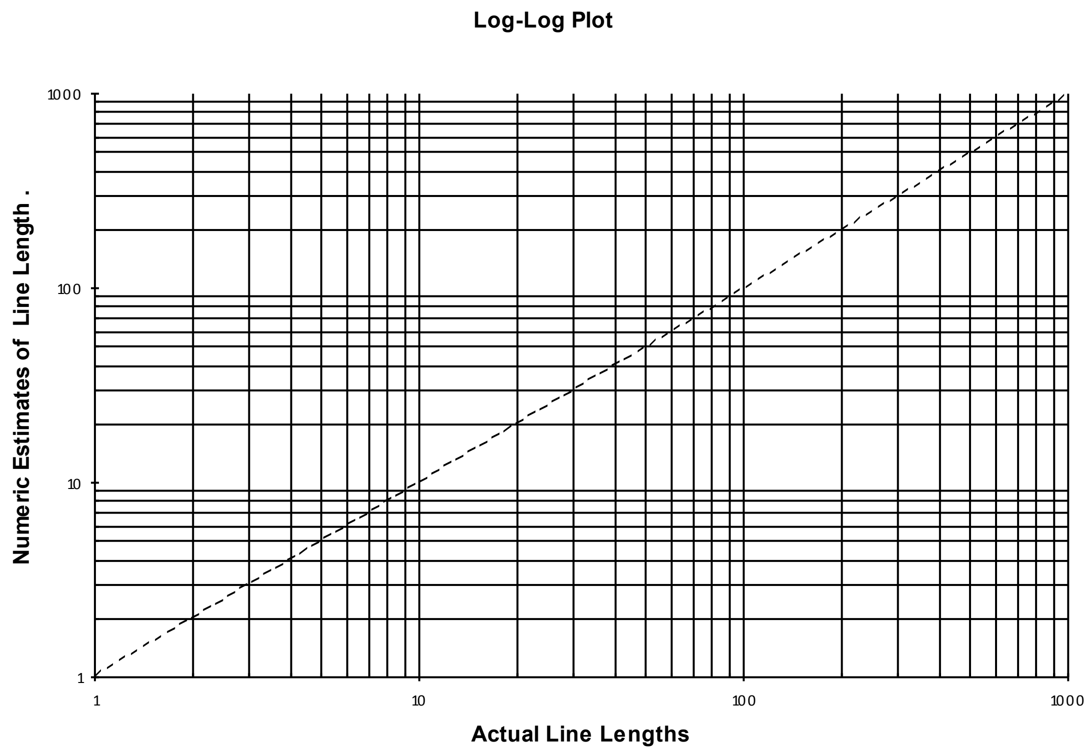
Compare your results to the actual line lengths of: 25, 79, 38, 150, 44, 127, 3, 110, 98, and 14. If we were to record the results from everyone in the room, we might be interested in seeing if, on average, our perceptions of line lengths are accurate. We're not interested in the actual line lengths (no one could tell us the reference line was 50 units long, because the units have no inherent meaning); we're interested in our perceptions of the line lengths relative to the reference line. If Line #2 is twice as long as Line #3, we want our average estimate of Line #2 to be twice as large as the average estimate of Line #3. Because of this, we need to calculate the geometric mean of our estimates.

Geometric Mean:
1. Take the logarithm of each subject's estimate for each line
2. Calculate the arithmetic mean of those log-values
3. Raise 10 to the power of the result obtained from the previous step.

$$\overline{X}_i = 10^{\left(\frac{\sum \log(x_{ai})}{n}\right)}$$

(with subject i's estimate for line a)

We then plot the average line length estimates against the actual line lengths on a log-log plot. When the axes of a graph are logarithmically ruled, equal distances on each axis mark off equal ratios. If we accurately perceive and estimate relative differences in line lengths, the resulting graph should be a straight line. Plot your results (or the class average) and see how close you came to the actual line length ratios.



Researchers in psychophysics have found the principle underlying this relationship – *equal stimulus ratios produce equal subjective ratios* – holds for many aspects of our five senses (not just our visual perceptions of line lengths). This principle is called the *Power Law*.

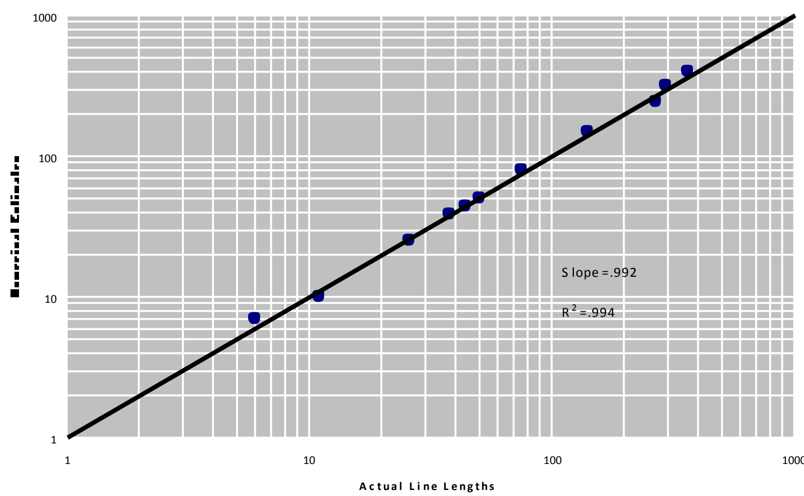
Examples of the Power Law are displayed on the following page…

Power Law: $R = kS^b$, where R = the subjective perception of the stimulus magnitude
S = the actual magnitude of the stimulus
k = a constant
b = the empirically derived exponent that characterizes the relationship (the characteristic exponent)

Restated: $\log(R) = b\log(S) + \log(k)$ or *The perceived (log of the) magnitude of the stimulus is a linear function of the actual (log of the) magnitude of the stimulus.*

If we graph the estimated magnitudes against the actual magnitudes on a log-log plot, we can find the value of b (the characteristic exponent or the slope of the line) through a simple linear regression. The following chart displays the results obtained from my 83 subjects. The chart provides evidence that the Power Law holds for visual estimates of line lengths. It also provides evidence that the characteristic exponent for visual perception of line length is 1.0.

Figure # 1: Perceived Line Lengths vs. Actual Line Lengths



Stimulus Modalities	Exponent	Description
Visual Length	1.00	Projected Line
Visual Area	0.70	Projected Square
Loudness	0.67	Sound pressure of 3000-Hz tone
Duration	1.10	White noise stimuli
Electric Shock	3.50	Current through fingers
Heaviness	1.45	Lifted weights
Muscle Force	1.70	Hand-grip contractions
Warmth	1.60	Metal contact on arm
Cold	1.00	Metal contact on arm
Smell	0.60	Heptane
Taste	1.40	Salt
Taste	1.30	Sucrose
Lightness	1.20	Reflectance of gray papers

Researchers have conducted studies to derive the characteristic exponents for other psychophysical domains. These exponents show the proportional increase in perceived stimulus magnitude when the actual magnitude has been doubled. For example, an exponent of 1.0 means that when the magnitude of a stimulus is doubled, subjects perceive a doubling of the magnitude of the stimulus.

The exponent for loudness is 0.67. This means that when the volume of a sound is doubled, subjects perceive only a two-thirds increase in the volume.

Subjects were then asked to draw 10 lines of varying lengths relative to a reference line (a line production exercise).

Reference Line **Length = 50**

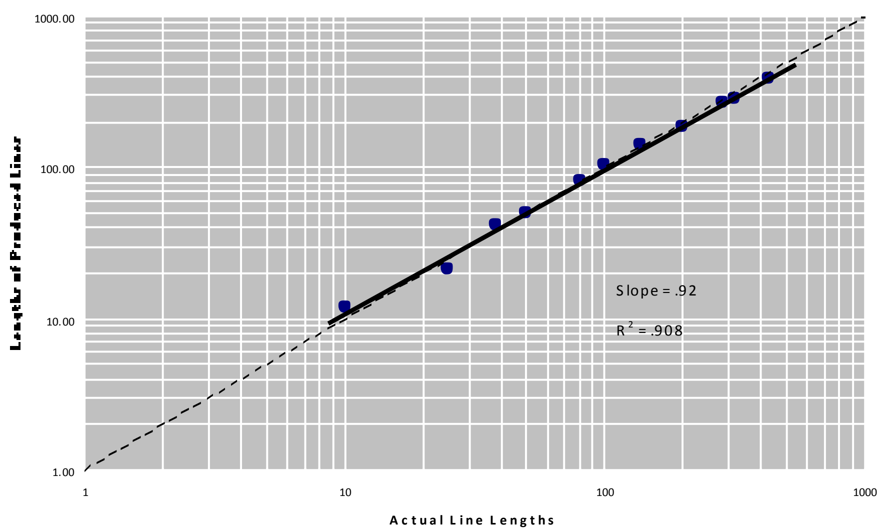
Line #1 **Length = 100**

Line #2 **Length = 20**

Line #3 **Length = 80**

After measuring the lines produced by each subject, the geometric mean length was calculated for each line. These geometric means were then plotted. Once again, the Power Law held.

Figure #2: Line Production Exercise



This provided evidence that the Power Law does hold in the psychophysical domain. In psychophysics, the horizontal axis will always be well-defined (with agreed-upon units of measurement). This makes it easy to check the validity of the Power Law for physical stimuli.

Social/psychological stimuli (such as attitudes) do not have an agreed-upon metric. For example, if I'm interested in measuring the hostility countries have towards the U.S., I cannot put the actual level of hostility on the horizontal axis (no unit of measurement exists). How can we demonstrate the Power Law applies for social data? What do we put on the horizontal axis?

Suppose we obtain estimates of line lengths in two different ways.

1. Suppose we first have subjects squeeze a hand-grip to estimate the length of each line. If a subject believes a line is twice as long as the reference line, the subject would attempt to squeeze the grip twice as hard.

2. Suppose we then show the subjects a reference line and play a tone. Subjects would then adjust the volume of the tone to match the perceived increase in stimulus magnitude. Thus, if a subject believed a line was three times as long as the reference line, the subject would attempt to adjust the tone so that it becomes three times louder.

Cross-Modality Matching

If the Power Law is valid and
if the exponents derived from magnitude estimation are truly characteristic (the exponents listed on page 5 are invariant)
then any two response measures with established exponents could be used to estimate the magnitude of stimuli.

In our situation: Hand-Grip: $R_1 = kS_1^{1.70}$ and Loudness: $R_1 = kS_1^{1.70}$

$$(1.70)\log(S_1) = (0.67)\log(S_2)$$

$$\log(S_1) = \frac{0.67}{1.70}\log(S_2) = 0.394\log(S_2)$$

General Form: $$\log(S_1) = \frac{b_2}{b_1}\log(S_2)$$

Thus, if we were to plot the two estimates (hand-grip and loudness) of line lengths against each other on a log-log graph, it should result in a line with a slope equal to the ratio of the characteristic exponents. When we do this, we don't need a variable with an agreed-upon metric on the horizontal axis.

Steps to demonstrate validity of Magnitude Scales:

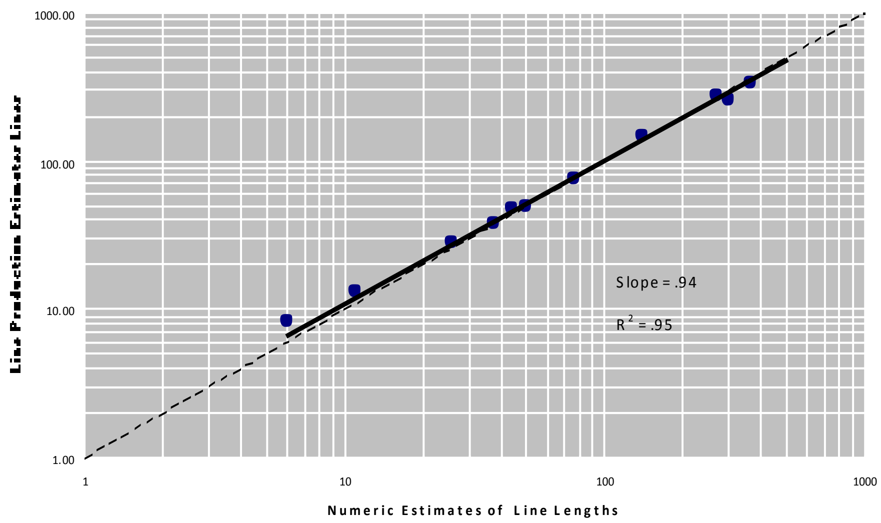
1. **Calibration Exercise**
   a) Have subjects numerically estimate the lengths of a set of lines relative to a reference line. Calculate the geometric mean length for each line.
   Plotting the estimates against the actual lengths on a log-log plot should result in a line with a slope of 1.0. This shows subjects understand ratio relationships.
   b) Have subjects draw lines of varying lengths relative to a reference line. Once again, the resulting line on a log-log plot should have a slope of 1.0
   c) Plot the two estimates from a & b against each other. The resulting line on a log-log graph should have a slope equal to the ratio of their exponents (1.0 / 1.0 = 1.0)
   This provides criterion-validity evidence of the magnitude estimates.

2. **Data Collection**
   a) Have subjects provide numerical estimates of the magnitude of a set of stimuli (20 countries and their hostility towards the U.S.) relative to a reference stimuli. Calculate geometric means for each stimulus.
   b) Have subjects draw lines corresponding to their perceptions of the magnitude of the same set of stimuli. Calculate the geometric means for each stimulus.
   c) Plot the two estimates from a & b against each other. The resulting line on a log-log graph should have a slope equal to the ratio of their exponents (1.0 / 1.0 = 1.0)

3. **Other possible steps**
   a) Lodge (1981) provides steps to scale subjects instead of stimuli
   b) Analyze individual responses to identify subjects who do not understand the magnitude scaling process (quite a few of my algebra students)
   c) Take the arithmetic average of the numerical estimates and line production estimates to come up with the final scale values
   d) If no zero point was specified, the zero is arbitrary.

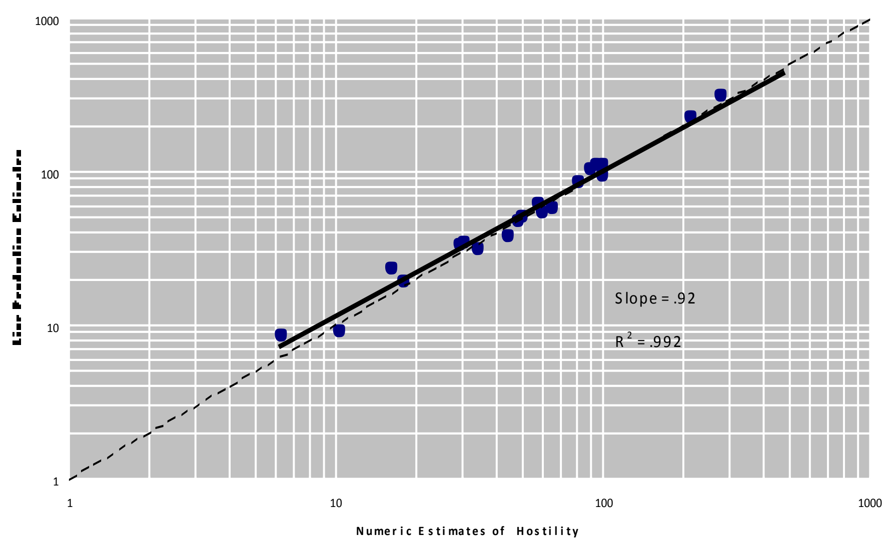
Cross-Modality Evidence for the line length estimates:

Figure #3: Cross-Modality Matching



Cross-Modality Evidence for the hostility estimates

Figure #4: Perceived Hostility Towards the U.S.



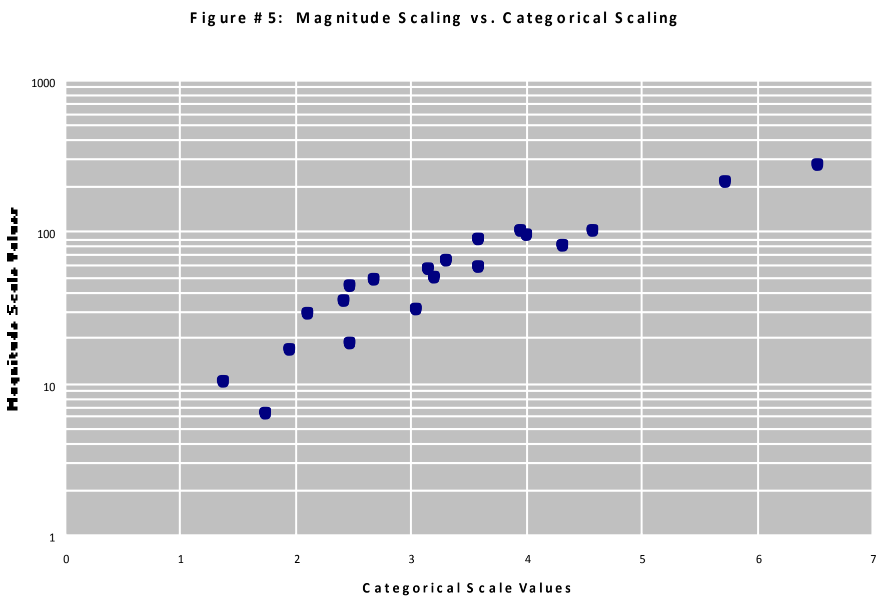
Magnitude Scaling Advantages:

1. Ratio level of measurement (with an arbitrary zero point)
2. Category labels do not influence subject responses

Magnitude Scaling Disadvantages:

1. Time consuming – the survey took at least 10 minutes to complete compared to about 2 minutes for the categorical survey
2. Analyzing the data is a bit more difficult – measuring lines, calculating geometric means, and plotting results on logarithmically scaled axes can be difficult
3. Quite a few mathematically-challenged students didn't understand the task.

In order to eliminate these disadvantages, we could assume the Power Law holds and have subjects only provide numerical estimates (skip the calibration and line production exercises)



	Standardized		
Country	**Numeric**	**Line Production**	**Categorical**
Iraq	3.07	3.17	2.47
Afghanistan	2.12	1.99	1.87
Saudi Arabia	0.42	0.25	0.49
Pakistan	0.41	0.48	0.98
Cuba	0.34	0.48	0.53
China	0.27	0.36	0.21
Bosnia	0.15	0.13	0.77
Russia	-0.09	-0.24	0.00
Israel	-0.18	-0.28	0.21
Germany	-0.21	-0.17	-0.12
France	-0.32	-0.34	-0.08
Japan	-0.34	-0.37	-0.48
India	-0.40	-0.50	-0.64
Panama	-0.55	-0.59	-0.68
Spain	-0.60	-0.55	-0.20
Mexico	-0.62	-0.57	-0.93
Canada	-0.79	-0.76	-0.64
Great Britain	-0.81	-0.71	-1.05
Sweden	-0.90	-0.90	-1.49
Australia	-0.96	-0.90	-1.21