THE ROLE OF PERCEPTION IN STATISTICS-BASED DECISIONS

Bryan Fricke

Statistical Discovery.™ From SAS.

SAS Campus Drive
Building T
Cary, NC 27513
1.919.677.8000

Contact Information

Bryan Fricke, JMP Senior Software Developer, SAS, Bryan.Fricke@jmp.com

Abstract

JMP is a powerful tool for generating statistical reports for evaluation by decision makers. However, when it comes to preparing reports, accuracy and comprehensibility are only part of the story. For example, Amos Tversky and Daniel Kahneman have suggested that presenting results in terms of a potential loss can have about twice the psychological impact as an equivalent gain. In this session, we will explore the role perception plays in statistics-based decisions and how knowledge of that role should inform JMP users with respect to generating reports for decision makers.

Introduction

After you have conducted an exploratory data analysis (EDA) and made discoveries you want to share, you need to determine how to communicate your message effectively. Creating an effective involves conveying information quickly and without ambiguity [1]. While doing so may seem straightforward, there are more twists and turns than one might expect. For example, if you provide unambiguous information about the probability of various options, are you quite sure you know how your audience will perceive that information? In the book, *Thinking, Fast and Slow*, Daniel Kahneman details how our automatic and unconscious perceptions can influence our decisions.

Expected Utility Theory

For a statistician, the value of a risky (probabilistic) prospect is its expected value. Mathematically, expected value is the sum of each outcome weighted by its probability.

$$E[v(x)] = \sum_{i=1}^{n} p_i x_i$$

Where,

- $E[v(x)]$ is the expected value of outcomes
- $v(x)$ is value as a function of x
- $x_1, x_2, \ldots, x_n$ are expected outcomes
- $p_1, p_2, \ldots, p_n$ are probabilities for each outcome

So, for example, a 25% chance to win $500 and 75% chance to win $100 is computed as follows:

$$E[v(x)] = 0.25 \times \$500 + 0.75 \times \$100 = \$200$$

Nonetheless, common preferences involving choices between simple gambles often contradict the expected value criterion. Take, for example, the following two options.

1. 80% chance to win $10M and 20% chance to win $1M dollars

   OR

2. $8M for sure

The expected value of the first gamble is $8.2M, while the expected value of the second is $8M. Even so, most people prefer the second gamble [2].

In 1738, a Dutch mathematician and physicist named Daniel Bernoulli developed the expected utility hypothesis, which explains people's general aversion to risk. Bernoulli suggested people attempt to optimize a hidden variable named "utility" rather than the expected monetary value of a gamble. He explained his formulation of utility by suggesting a small change in utility is inversely proportional to a person's current wealth. From that basis, he showed that utility is a logarithmic function of wealth [3].

$$du(w) = k\frac{dw}{w}$$

$$u(w) = k\ln w + C$$

$$C = -k\ln w_0$$

$$u(w) = k\ln\frac{w}{w_0}$$

Where,

- $u(w)$ is utility as a function of wealth
- w is current wealth
- w_0 is initial wealth
- k and C are both constants

Figure 1 shows the logarithmic relationship between wealth and utility.

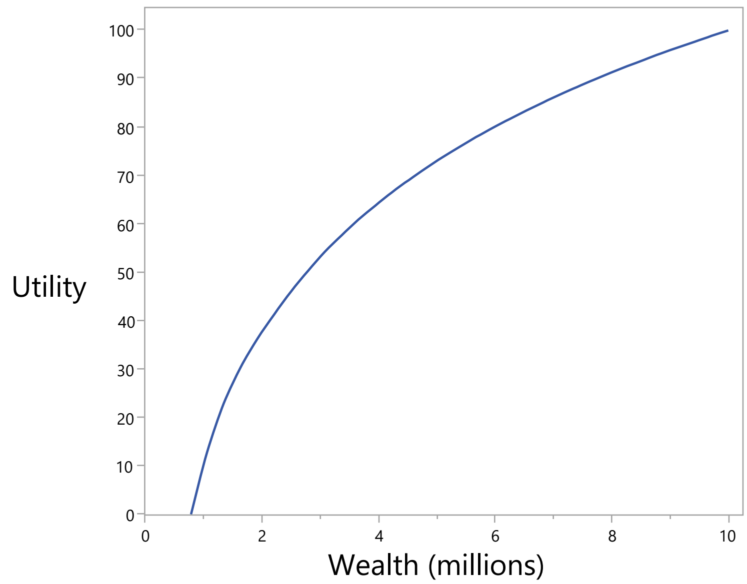

FIGURE 1. LOGARITHMIC RELATIONSHIP BETWEEN WEALTH AND UTILITY

So, rather than weighting each outcome's monetary value, Bernoulli posited that people weight the utility of each outcome when assessing the value of a gamble.

$$E[u(w)] = \sum_{i=1}^{n} p_i k \ln \frac{w_i}{w_0}$$

Where,

- $E[u(w)]$ is expected utility
- $u(w)$ is utility as a function of wealth
- $w_1, w_2, \ldots, w_n$ are expected wealth outcomes
- $p_1, p_2, \ldots, p_n$ are probabilities for each outcome
- w_0 is initial wealth

With the above formulation, we can evaluate the expected value of the gamble presented earlier. If we assume that k is 38.95 and w_0 = $773K, we have the following:

$$E[u(x)] = 0.8 \times 100 + 0.2 \times 10 = 82 \rightarrow \$6.3e^6$$
$$E[u(x)] = 1.0 \times 91 = 91 \rightarrow \$8.0e^6$$

Figure 2 shows where these values appear on the graph of utility versus wealth.

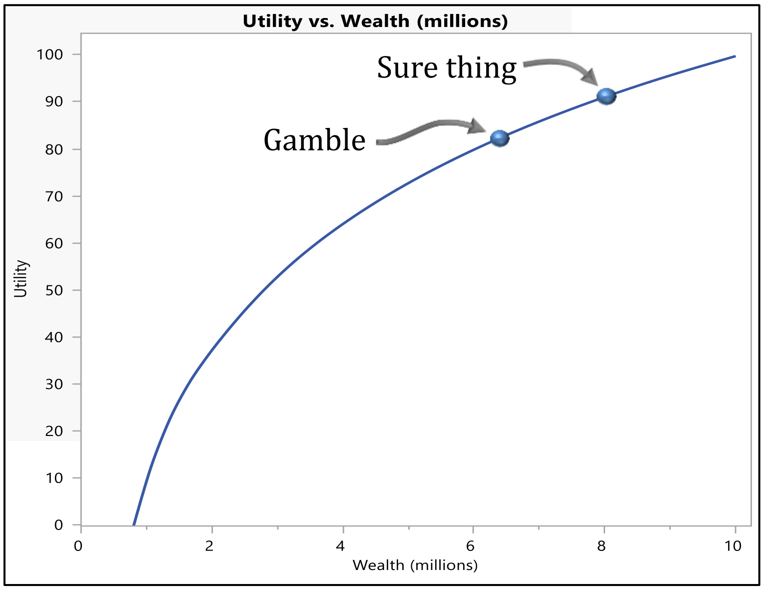

FIGURE 2. EXPECTED UTILITY FOR A RISKY PROSPECT AND A SURE THING

In this case, expected utility theory correctly predicts the sure thing will be the more popular choice even though it has the lower expected value. Because of the concavity of the curve, people are expected to be risk averse. Thus, for equal chances for an equivalent gain or loss, utility theory predicts the utility associated with the loss will be greater than that of the gain [2].

In 1947, John von Neumann, a Hungarian-American mathematician, physicist, and computer scientist, and Oskar Morgenstern, a German-born economist, formally proved that a "rational" person faced with risky choice will act to maximize the expected value of a function defined over the domain of outcomes, such as Bernoulli's. To be rational, the actor needs only to satisfy four axioms of rational behavior as defined below.

1. Completeness: $A \succ B \text{ or } A \sim B \text{ or } B \succ A$

   A person has well-defined preferences between any two options.

2. Transitivity: $If\ A \succ B\ and\ B \succ C, then\ A \succ C$

   A person has consistent preferences.

3. Independence of alternatives: $A \succ B\ if\ and\ only\ if\ ApC \succ BpC$

   A person's preference between two gambles is independent of the presence of a third gamble.

4. Continuity: $If\ A \succ B \succ C\ there\ exists\ some\ p\ such\ that\ ApC \succ B \succ A(1-p)C$

   A person that prefers gamble A to B and B to C will be equally happy with B or a combination of A and C for some unique probability, p, such that A has probability p and C has probability $1-p$.

Of the four, the first and second are most salient. The third is often relaxed when objections are raised about the validity of the axioms. The fourth is introduced mainly for mathematical tractability. A person that satisfies all four conditions is said to be von Neuman-Morgenstern (vNM) rational [4].

Richard Thaler, an American economist, distinguishes between theoretical “econs” and “humans”. The notion of an econ comes from the economics discipline in which it is assumed that people are vNM rational, self-interested and have unchanging preferences. On the contrary, humans have bounded rationality (i.e., they have limited cognitive abilities), limited self-interest, and their preferences are context dependent [2].

Prospect Theory

In 1979, Daniel Kahneman and Amos Tversky, two Israeli psychologists, observed that people are risk averse with respect to gains, as Bernoulli suggested; however, they are risk seeking with respect to losses. For example, which gamble would you choose in each of the following cases?

1. Gain $900 for sure

   OR

   98% chance to get $1,000

2. Lose $900 for sure

   OR

   98% chance to lose $1,000

Experiments show most people prefer the sure thing in the first case, and the gamble in the second case. Bernoulli’s logarithmic function cannot explain why a person would be risk seeking in the case of losses [2].

Frames of Reference

In addition, Kahneman and Tversky observed being risk averse or risk seeking in the above case is independent of wealth. Rather, a loss or gain is with respect to a perceived reference point, which often is the person’s status quo or adaption level. The notion of adaption level is taken from adaption-level theory which was development in 1947 by Harry Helson, an American psychologist. Adaption-level theory suggests that perception of a stimulus depends on the current stimulus state relative to a subject’s current adaption level. The current adaption level is based on previous experiences with the stimuli [5].

Helson incorporated adaption level into a reworking of Fechner’s law of perception. Fechner’s law was developed in 1860 by Gustav Fechner, a German philosopher and physicist. Interestingly, Fechner’s Law was developed in the same way that Bernoulli developed his utility hypothesis over a hundred years earlier. First, Fechner formulated Weber’s contrast [6].

$$dp = k\frac{dS}{S}$$

Where,

- p is the current level of perception
- S is the current stimulus state

- k is a constant

Weber's contrast expresses the inverse relationship between a just noticeable difference and the physical output of a stimulus.

From there, Fechner developed Fechner's law which shows that perception is a logarithmic function of the current state of the stimulus relative to some initial stimulus state.

$$p = k \ln S + C$$

$$C = -k \ln S_o$$

$$p = k \ln \frac{S}{S_0}$$

Where,

- S_0 is initial stimulus state

For Fechner, the initial stimulus state is the state to which the person a person is unable to detect the stimulus. Like Kahneman and Tversky, Helson believed the denominator should be a person's current adaption level.

To get an idea of adaption level, think of three bowls of water as shown in Figure 3.

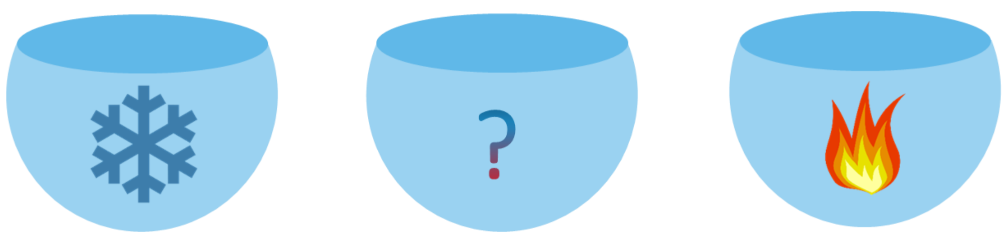

FIGURE 3. SETUP FOR ADAPTION LEVEL EXPERIMENT

One bowl is filled with cold water, the middle bowl is filled with water at room temperature, and the third is filled with hot water. Now imagine you place your left hand in the bowl with cold water and your right hand in the bowl with hot water for a minute. Afterwards, you put both hands in middle bowl. As might imagine, the water will feel warm to your left hand and cool to your right hand [2].

Likewise, Kahneman and Tversky suggested losses and gains from uncertain or risky prospects are perceived relative to a person's current adaption level or frame of reference. A person's current frame of reference is heavily influenced by what information is readily available. Consequently, a person's reference level can be influenced by the way a question is framed as will be explained later. The idea that choice can be affected by presenting the same option in different ways challenges the vNM rational axioms of completeness and transitivity [2].

Loss Aversion

Kahneman and Tversky also observed that losses have about twice the psychological impact of an equivalent gains. For example, consider the following gamble:

50% chance to lose $100

AND

50% chance to win $150

If you are like most people, you would not accept the gamble although it has a positive expected value. You can measure your own aversion to loss by determining the amount of gain you would require to offset a loss with equal odds. Experiments show a gain would generally need to be between 1.5 and 2.5 times as large as the loss. That said, those who are in a profession that routinely trades commodities are generally less loss averse. Kahneman reports that loss aversion can be partially overcome by being told to "think like a trader" [2].

The Three Principles of Prospect Theory

Putting the ideas of reference frames and loss aversion together with Bernoulli's idea of diminishing sensitivity to changes as values increase, Kahneman and Tversky developed Prospect Theory which consists of three principles [2].

1. Psychological value is relative to neutral reference point (adaption level) rather than wealth
2. Decreasing sensitivity to changes as amounts increase
3. Losses are more aversive than gains are attractive

These three concepts are illustrated graphically in Figure 4.

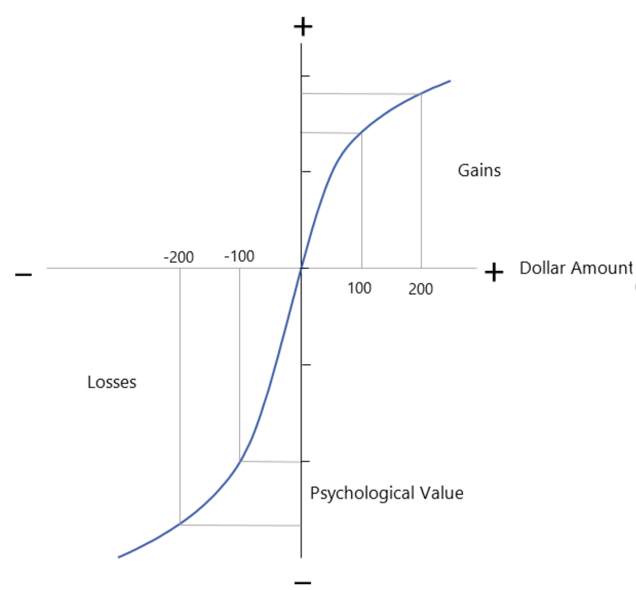

FIGURE 4. GRAPHICAL REPRESENTATION OF PROSPECT THEORY

Decision Weights

In addition to the three principles of Prospect Theory, Kahneman and Tversky observed that decision weights are not proportional to probability. For example, do you perceive the following changes as equal improvements?

- From 0 to 5%
- From 5% to 10%
- 60% to 65%
- 95% to 100%

If you are like most people, the changes from 0 to 5% and 95 to 100% are more significant than the other changes in probability. The change from 0 to 5% tends to create a sense of hope or possibility that is weighted disproportionately to the change in odds. This disproportionate weighting of small odds is known as the possibility effect. Similarly, a move from 95 to 100% creates a sense of certainty that is missing from a 95% chance. In this case, the chance of losing is overweighted. The underweighting of highly favorable odds compared to a sure thing is known as the certainty effect.

Kahneman used experimental results to create a mapping of the psychological weight to probability as shown in Figure 5.

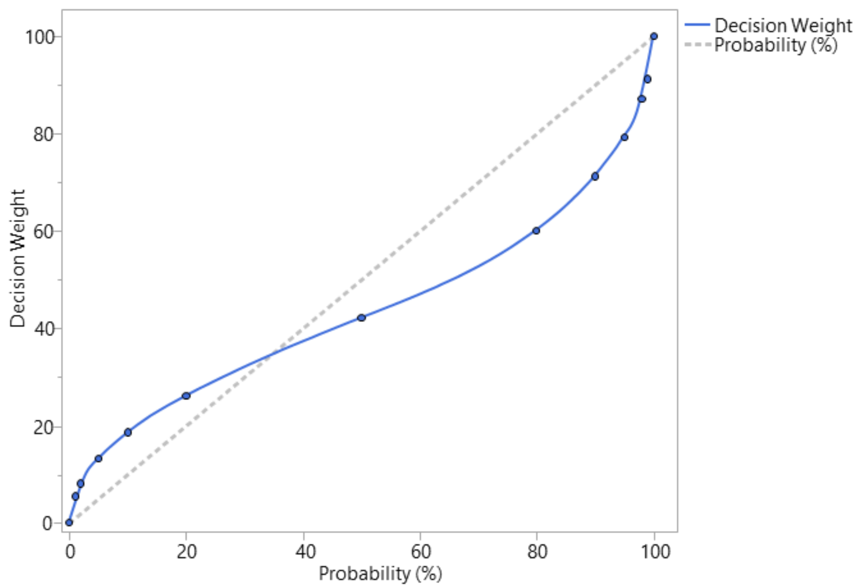

FIGURE 5. MAPPING OF DECISION WEIGHTS TO PROBABILITIES

As can be seen in Figure 5, these decision weights are not linear with probabilities. Kahneman also adds that if small probabilities are not overweighted, they are neglected. Von Neumann and Morgenstern proved that a weighting function that is not proportional to probability will result in decisions that lead "to inconsistencies and other disasters [2]." In other words, if decision weights are not a linear function of probability, they do not qualify as vNM rational.

The Fourfold Pattern

Put decision weights together with the three principles of prospect theory and you get the following fourfold pattern.

	GAINS	LOSSES
HIGH PROBABILITY Certainty Effect	95% chance to win $10,000 Fear of disappointment RISK AVERSE	95% chance to lose $10,000 Hope to avoid loss RISK SEEKING
LOW PROBABILITY Possibility Effect	5% chance to win $10,000 Hope of large gain RISK SEEKING	5% chance to lose $10,000 Fear of large loss RISK AVERSE

In the top row, both decreasing sensitivity and underweighting of favorable probabilities contributes to risk aversion for gains and risk seeking for losses. Decreasing sensitivity operates in the same fashion in the bottom room; however, the overweighting of low probabilities is more significant than the curvature of the value function. As such, people become risk seeking for gains and risk averse for losses when the probabilities are small [2].

The upper left cell suggests why people are willing to accepted “structured settlements” that offer less than expected value when involved in a lawsuit they are likely to win. The bottom left cell explains why lottery tickets are popular. The bottom right cell explains the attraction of insurance. The upper right cell explains the sunk cost fallacy where people throw good money after bad in an attempt salvage lost causes [2].

Endowment Effect

As mentioned, professional traders show less loss aversion than others and being told to “think like a trader” reduces loss aversion. In these cases, the objects of outcomes are for trade rather than use. So, when items are held for consumption rather than trade, aversion to loss is more pronounced. This effect is called the endowment effect. For example, in one experiment one group of participants, called Sellers, were given customized coffee mugs and asked how much money it would take for them to sell them. Another group of participants, known as Buyers, were asked how much they would be willing to pay for one of the mugs. Finally, a group of participants, called Choosers, had the option to receive either a coffee mug or an amount of money they thought would be equal in value. The average price that each group placed on the coffee mug is as follows [2]:

Sellers

$7.12

Choosers

$3.12

Buyers

$2.87

Notice that Sellers asked over twice as much for the coffee mugs as either Choosers or Buyers. Also, note that Sellers and Choosers faced the same choice. The only difference was in how the choice was framed. For the sellers, they possessed the mug prior to determining its value [2].

Denominator Neglect

Decision weights can be further warped by a phenomenon known as denominator neglect. Denominator neglect can occur when ratios are treated as fractions rather than percentages. When treated as fractions, numerators call out a discrete number of individual items which can be more vividly imagined. Thus, the numerator becomes the focus of attention and the denominator is neglected. For example, which urn would you select from if a red marble wins a prize?

- Urn A contains 10 marbles, of which 1 is red.
- Urn B contains 100 marbles, of which 8 are red.

Drawing from urn A offers a higher probability of winning a prize and yet 30 to 40% of students participating in such an experiment chose the urn with the greater number of red marbles [2].

Reframing

As mentioned earlier, the frame of reference can influence a person's decision. As discussed, people are about twice as averse to losses as they are attracted to equal gains. Thus, loss aversion creates something like a gravitational field around the status quo. To overcome loss aversion, a gain can be framed as a potential loss. Framing an option in this way may involve having the person imagine they have already taken the option with the higher expected value and are now viewing the loss that would have resulted with sticking with the status quo. For example,

Frame 1: You have been given $1,000. Now which option would you choose?

50% chance to gain $1,000 OR gain $500 for sure

Frame 2: You have been given $2000. Now which option would you choose?

50% chance to lose $1000 OR lose $500 for sure

Most people choose the sure thing in Frame 1 and the gamble in Frame 2 [1].

Applications

With Graph Builder, JMP makes it easy to create a variety of chart types to convey what you have learned after an Exploratory Data Analysis (EDA). If you are not careful; however, you will generate a graph or report that is perfect for econs, but less so for humans. Let's go through some applications of Prospect Theory to reports and graphs.

Establish a Baseline

When creating a graph, JMP automatically scales axes so that data is readily visible. Even so, remember, if you are showing absolute values, users will have diminishing sensitivity to changes as values increase. As mentioned previously, people have a reference point or adaption level from which both gains, and losses are assessed. A starting point for evaluating both gains and losses is commonly referred to as a baseline. So, rather than showing absolute values, consider showing differences relative to a baseline. Otherwise, you leave it up to your audience to perform their own mental math.

Also, be careful not to set your baseline too low. The most obvious thing may be to use current performance as a baseline. If you are measuring a system that is underperforming, it may be better to use an industry average or a desired goal as the baseline. That way current performance will be shown as a loss compared to the baseline. Due to loss aversion, the motivation to recover from a loss will be roughly twice as much as if the same difference was described as a gain.

For example, Figure 6 shows a report taken from the EDA module in the free online JMP training course named *Statistical Thinking for Industrial Problem Solving*.

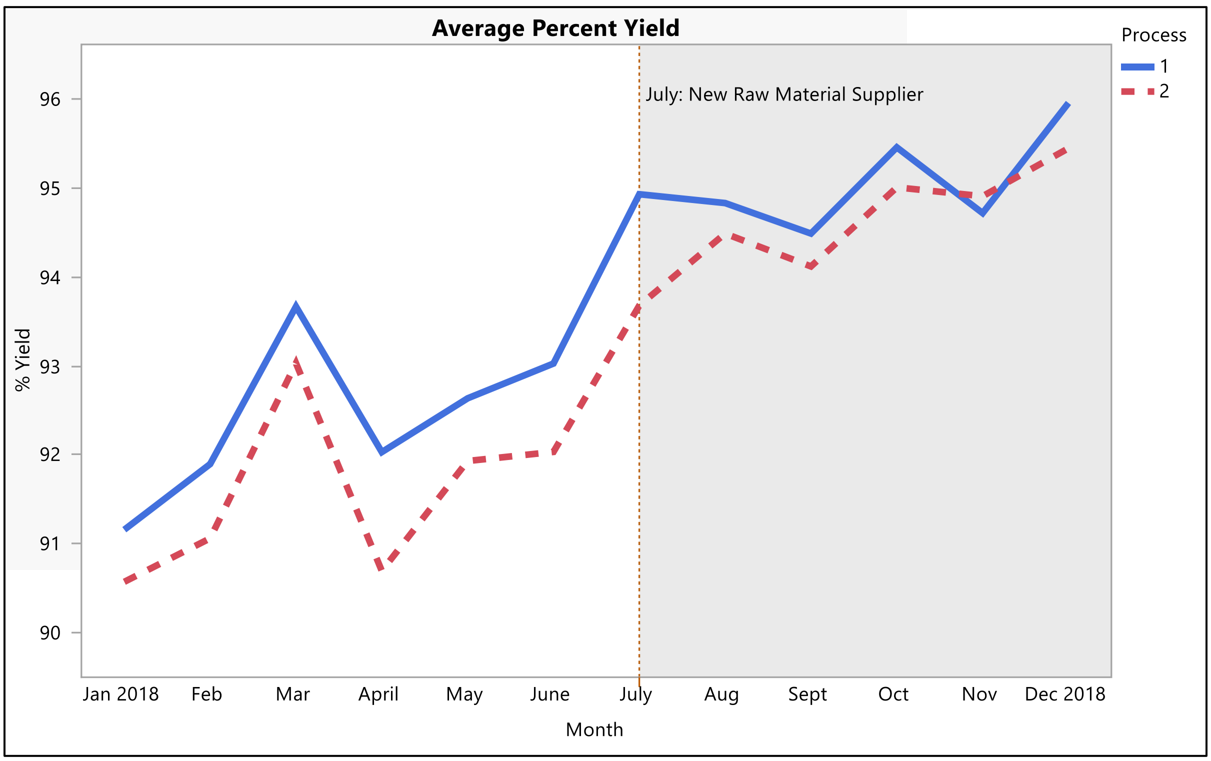

FIGURE 6. GRAPH OF PERCENT YIELD FOR PROCESS 1 AND PROCESS 2

The graph does a good job showing upward trends in percent yield for both Process 1 and Process 2. The graph also shows an apparent positive correlation with an increase in percent yield when a new supplier was introduced in in July. However, what if the question is, “Which process should we use?” In that case, we are looking at the differences between the two processes, which clearly favor Process 1 in this case. Nonetheless, it is difficult to see exactly how much of a difference there is between the two processes. The most obvious way to graph the differences may be to show the positive differences of Process 1 as compared to Process 2. Instead, we can graph the differences between both processes and a baseline. In the absence of a baseline, we can use Process 1 as the baseline and show most of the differences as losses as in Figure 7.

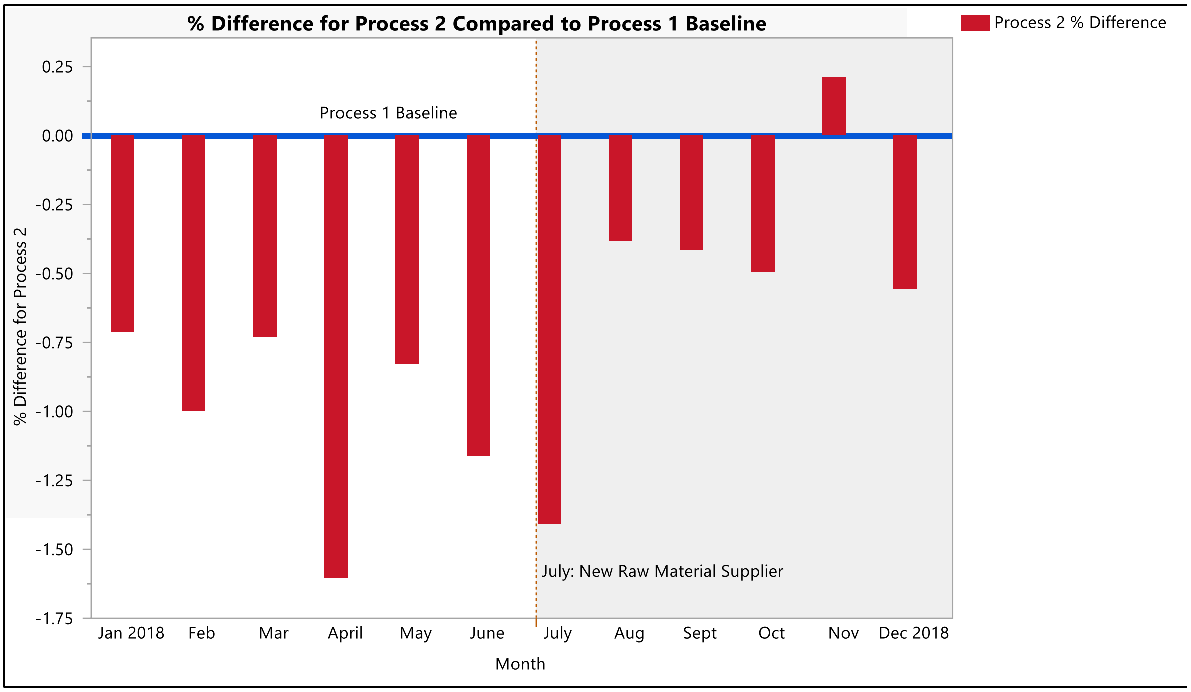

FIGURE 7. PROCESS 2 DIFFERENCES FROM PROCESS 1 BASELINE

Now we can more clearly see the values of the differences. In addition, we may be able estimate the average loss per month, which is -0.757%. Experiments show people effortlessly generate accurate estimates of the average of positive line lengths in a fraction of a second. Nonetheless, people do not have same ability to estimate summations. If we think our audience wants to see the sum for the entire year and as well as the trend of gain or loss, we may want to show a graph that shows a running total as in Figure 8.

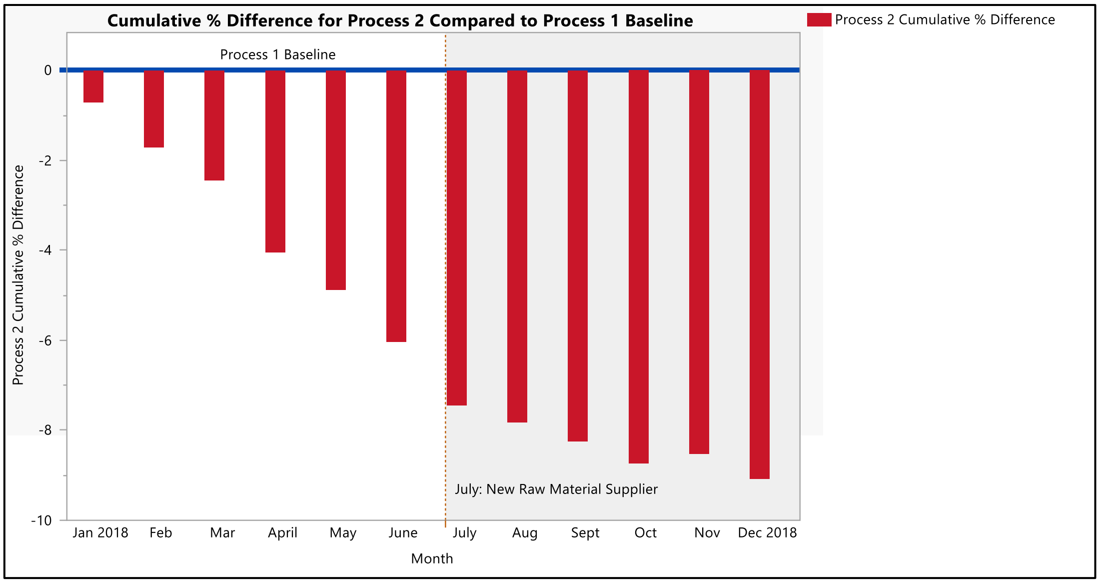

FIGURE 8. ACCUMULATION OF DIFFERENCES BETWEEN PROCESS 1 AND 2

If we conclude we only want to use Process 1, the question may be "should we use the old or new supplier?" With the original graph, we would need our audience to mentally transfer Process 1 values for the second half of the year over the values from the first half of the year and then perform mental subtractions. Instead we can do the work for them as shown in Figure 9.

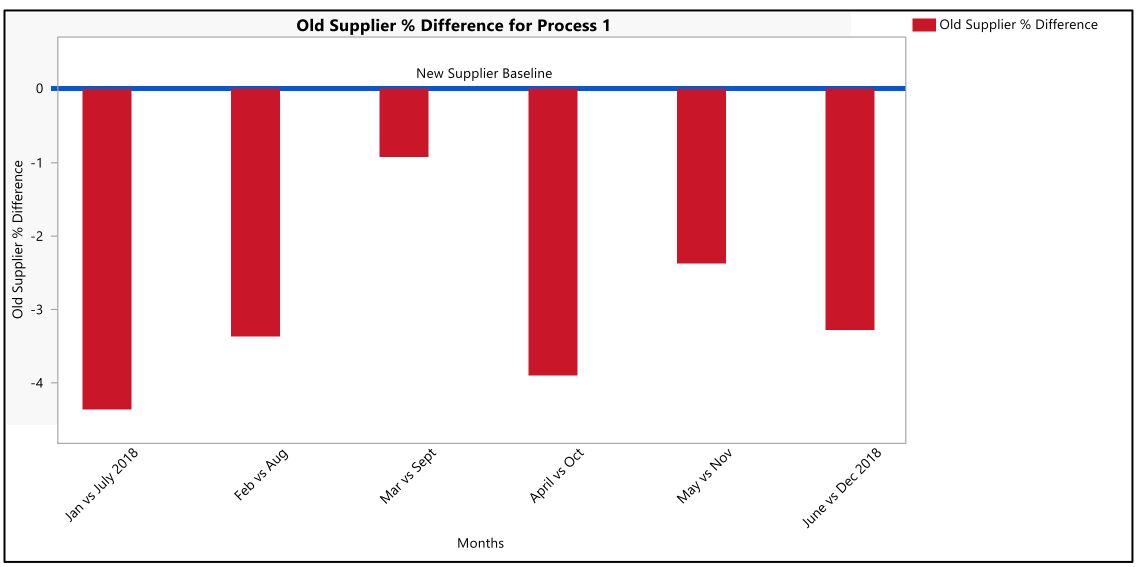

FIGURE 9. PROCESS 1 DIFFERENCES BETWEEN OLD AND NEW SUPPLIER

If instead, we want to show both the monthly trend of losses as well as the accumulation of losses for the year we can create a graph like the one shown in Figure 10.

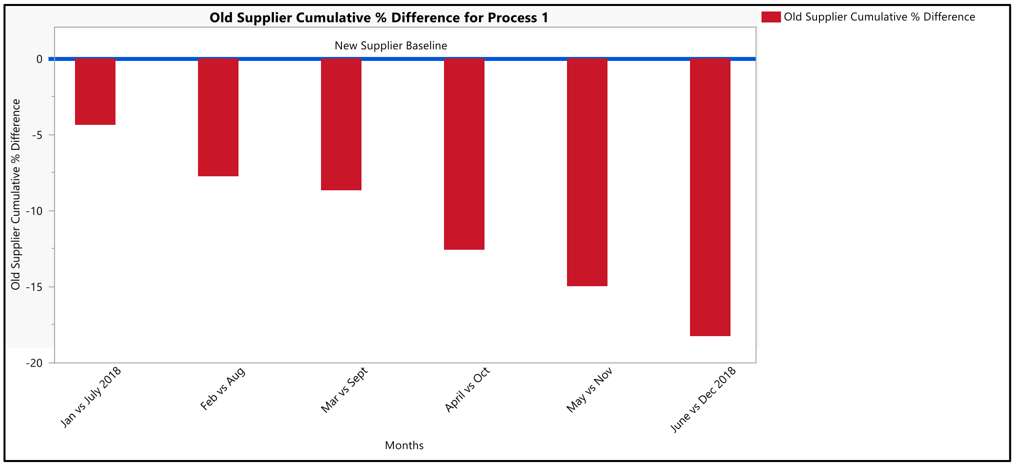

FIGURE 10. ACCUMULATION OF PROCESS 1 DIFFERENCES BETWEEN OLD AND NEW SUPPLIER

Create an Appropriate Frame

As described earlier, the way a question is framed can affect the answer. So, we need to be careful we do not inadvertently introduce a distorted or a mentally taxing frame. For example, consider the following question [2]:

> Assuming both Tom and Kim drive 12K miles per year, who will save more gas by switching cars?
>
> - Tom switches from a 12 mpg car to a 14 mpg car
> - Kim switches from a 30 mpg to a 40 mpg car

Since gallons is inversely proportional to mpg, a fixed yearly mileage implies gallons of gas used becomes increasingly insensitive to improvements in mpg as mpg increases.

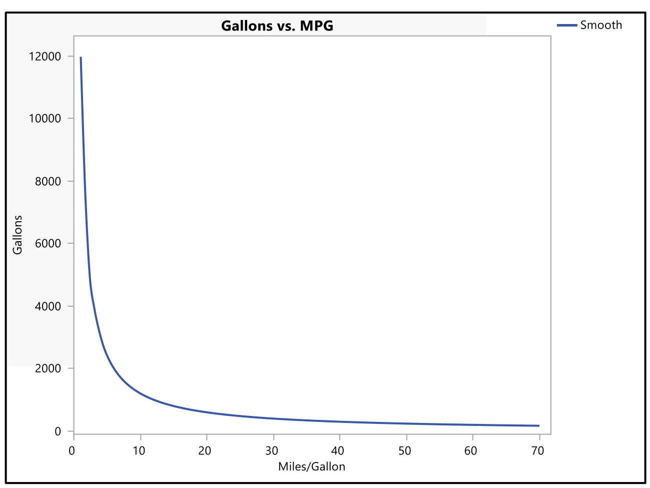

FIGURE 11

As it turns out, Tom would save 119 gallons whereas Kim would save 83 gallons. So, it would be better to consider ratios that are linear with the solution like gallons per 100 miles.

Likewise, if we are considering improving the percent yield of only one of the two processes shown earlier, we need to be careful we provide an appropriate ratio. If you can increase the yield of Process 1 from 92.402% to 95.066% (a difference of 2.664%) and Process 2 to from 90.553% to 93.195% (a difference of 2.642%), which process should you choose to improve?

Consider the definition of percent yield.

$$Percent\ Yield = \frac{Actual\ Yield}{Theoretical\ Yield} \times 100$$

For a fixed output, theoretical yield more closely predicts material cost than does actual yield. It makes more sense to compare the differences in the reciprocal of percent yield since theoretical yield is inversely proportional to percent yield as shown in Figure X.

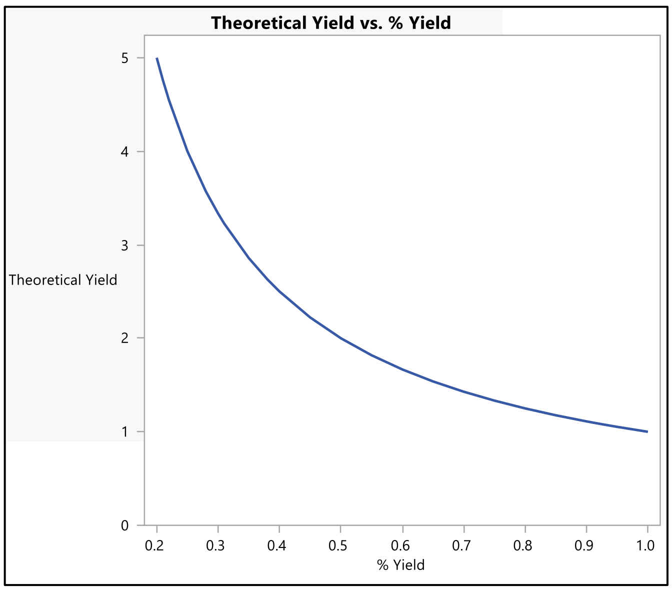

FIGURE 12. THEORETICAL YIELD IS INVERSELY PROPORTIONAL TO % YIELD

The change in the theoretical to actual yield ratio for Process 1 is 3.03% and 3.13% for Process 2. So, all other things being equal, we should improve Process 2 rather than Process 1.

Avoid Denominator Neglect

For graphs that show the values of small ratios, it may seem natural to present numerators with a denominator displayed somewhere on the report for context. Remember showing both numerator and denominator invites the viewer to focus on the numerator and neglect the denominator. In so doing, the viewer may have more concrete and vivid images of individual cases that effectively distorts the number of actual occurrences. Conversely, when presenting a value as a decimal, the viewer may neglect the values altogether since the numbers may appear so small and abstract. For example, the below graph attempts to address whether the measles vaccine has been effective across all states.

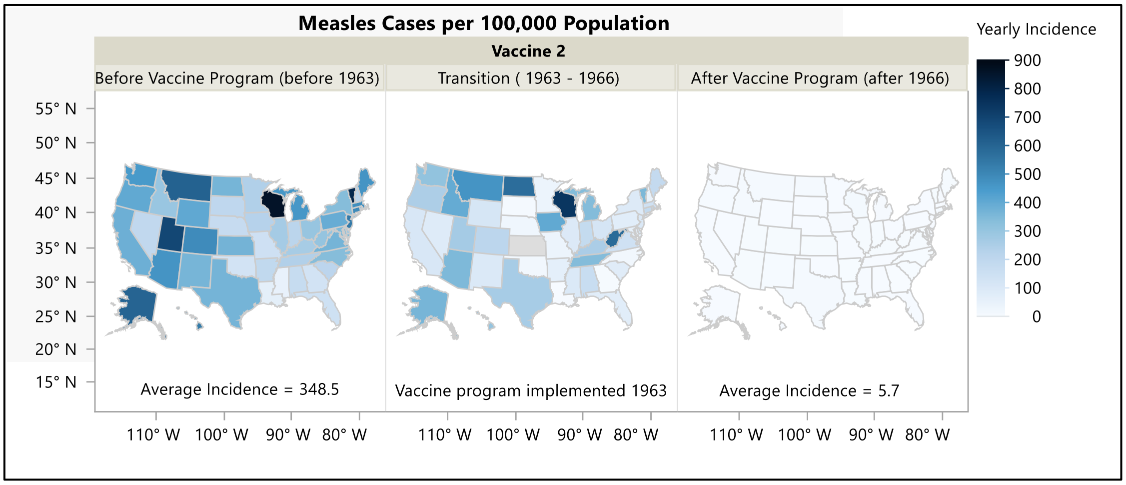

FIGURE 13. GRAPH SHOWING EFFECTIVENESS OF MEASLES VACCINE ACROSS ALL STATES

This graph uses both numerator and denominator to show the number of occurrences of measles. In the "Before Vaccine Program (before 1963)" graph, the average number of incidences across all states is shown to be 348.5 per 100,000 people. If this value had been shown as a percentage, it would have been show as 0.3485%. Such a value emphasizes the infection rate was well under 1% of the population before the vaccine was introduced. As discussed, when it comes to small percentages people will either overestimate its decision weight or neglect it altogether. In this case, showing the infection rate as a percentage may cause people to question the significance of the original infection rate. Presenting the value as 348.5 emphasizes a sizable number of cases of the disease for each state, which is likely the intent. The "After Vaccine Program (after 1966)" graph shows the average incidence rate to be 5.7 per 100,000 people. This value would have been shown as 0.0057% of the population, which means that, on average, under 1/100th of a percent of the population contracted the disease after the vaccine was introduced. The percentage is so small that many people may consider it negligible. Showing the value as "5.7" emphasizes the unfortunate individuals that still contracted the disease after the vaccine was widely available. That being the case, viewers may overestimate the scale of the problem after the vaccine was introduced. So, on one hand we want to emphasize the scale of the original problem and on the other we want to emphasize the low infection rate after the vaccine was introduced. Showing one value as a fraction and the other as percentage would be confusing at best and deceptive at worst. As it is, we can emphasize the individual cases in the "Before Vaccine" graph and the low infection rate in the "After Vaccine" graph by using 10,000 rather than a 100,000 as the denominator as shown in Figure 14. In that case, the average incidence would be shown as 34.9 in the "Before Vaccine" graph and 0.6 in the "After Vaccine" case. In this case, it is easier for the viewers to envisage individual cases before the vaccine and less so in the "After Vaccine" graph.

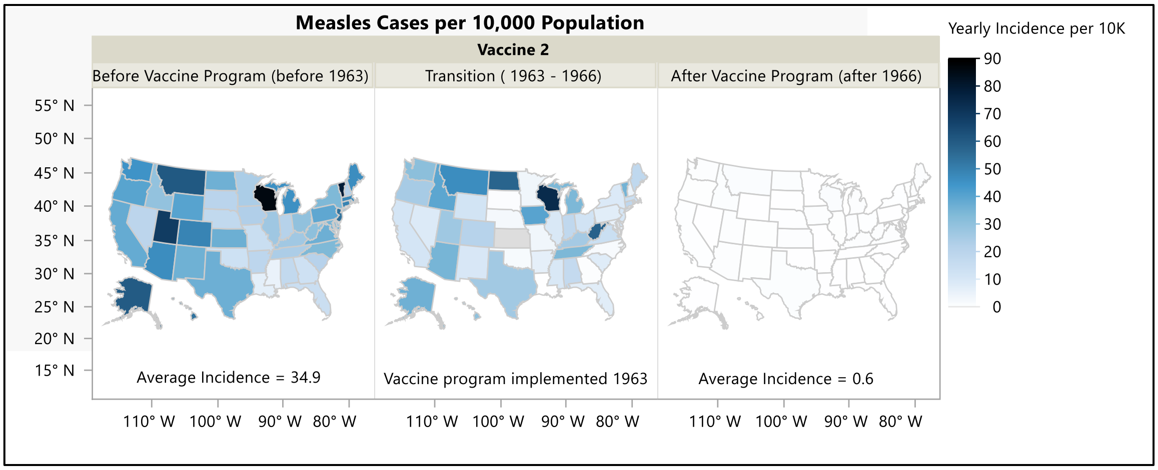

FIGURE 14. REVISED GRAPH SHOWING EFFECTIVENESS OF MEASLES VACCINE ACROSS ALL STATES

In Figure 14, also notice that the "After Vaccine" graph looks whiter than in the original graph. In the original graph, the scale was a light shade of blue at 0. Given the gradient was displayed against a gray background in the legend, the light blue looked whiter than it does against a white background which is a framing effect. Consequently, it would be difficult to distinguish 0 from 100 using the original gradient. In the latter case, it is easy to see that even the states with more cases than usual have, on average, 5 or fewer cases per 10,000 people.

Conclusion

JMP makes it easy to create graphs that are relevant, complete, and clear to any econ. The problem is we must communicate with humans that have limited processing abilities, biases, complex motives, and fickle preferences. Prospect Theory helps us understand how humans make decisions under uncertainty. The main takeaways are that people have the following characteristics that are relevant to generating reports and graphs.

1. Diminishing sensitivity to changes as values increase.
2. Gains and losses assessed from a reference point that can be influenced by the way a question is framed.
3. More aversive to losses than attracted to gains.
4. Tendency to overweight small probabilities and underweight large probabilities when making decisions.

Keep these characteristics in mind as you develop your reports, unless you are presenting to the fabled econs rather than mere mortals.

References

[1] S. Matange and D. Heath, *Statistical Graphics Procedures by Example: Effective Graphs Using SAS*. SAS Institute, 2011.

[2] D. Kahneman, *Thinking, Fast and Slow*, 1 edition. Farrar, Straus and Giroux, 2011.

[3] D. Bernoulli, "Exposition of a New Theory on the Measurement of Risk," *Econometrica*, vol. 22, no. 1, p. 23, Jan. 1954.

[4] M. Peterson, *An Introduction to Decision Theory*. Cambridge University Press, 2017.

[5] "Adaptation-level theory - Oxford Reference." [Online]. Available: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095350211. [Accessed: 20-Sep-2019].

[6] G. T. Fechner, *Elemente der psychophysik*. Leipzig : Breitkopf und Härtel, 1860.