

Data Availability: Influences of Data on Election Polling, Search Engines, Nutrition Labels, and Game Theory

By: Hwa Lang (Calvin) Cho

University of California, Irvine, School of Social Science, Ethics Center
2022 Internship Final Capstone Paper

Part I: Introduction

In October of 2012, Thomas H. Davenport and DJ Patil through the Harvard Business Review published the popularized article “Data Scientist: The Sexiest Job of the 21st Century.” The article was a reflection of a rapidly-growing sector of the technology industry: data science.

But what exactly is data science? Or more specifically, how does data science apply in our daily lives? Although the richness and complexities of data science are evidently seen, its daily application and uses aren’t always easily visible to the public eye. Data science, despite its position as the key player in many projects, usually serves as the stage crew who assist in operations backstage; this leads to many misconceptions and miscreditations to the data science behind many projects.

The purpose of this white paper will be to address the growing rise of data science and also its application in major fields such as election polling, search engines (SEOs), and nutrition labels (USDA).

Part II: Election Polling

I. Background and History

In 1948, Harry Truman ran for presidency following his impromptu “first-term” following the death of President Franklin D. Roosevelt as the Vice-President. Truman, the Democratic Party’s candidate, was up against New York Governor Thomas E. Dewey (R). Dewey presented many anti-crime policies and promised to curb the power of the American Mafia.

Toward the end of the campaigning season, many predictions indicated that Dewey would win the election as key electoral college states in New England showed strong support for Dewey. However, Truman’s campaign management had realized the growing electoral importance of the new western states and had launched a famous ‘whistle-stop tour’ of the country and successfully won the election with around 2 million more votes as well as winning the electoral count by 303-189. But because the eastern states had announced their decision first, the Chicago Tribune published an editorial with the banner headline “Dewey Defeats Truman,” incorrectly announcing the results of the election.

This was the first time that the utilization of data had seriously changed the fate of an election. It was evident that future elections would heavily rely on 1) polling and 2) strategies based off of those polls and data. However, polls have only gotten more and more inaccurate over the years. With no president exactly having a traditional “landslide victory” since Ronald Reagan in the 1980s, presidential elections have been ending with razor-thin marginal victories and defeats, making polls futile in predicting the fate of any election. Below is a case-by-case analysis of key presidential elections that resulted in marginal victories:

1) Election of 2000

During the Bush presidential campaign, the Mid-Year Gallup Report showed that the Bush campaign would likely win. In fact, the Gallup News Service had strong confidence in the fact that the 2000 election would end in Bush’s popular vote victory. Below is a statement by the service:

“In contrast, the 2000 election has been markedly consistent. Although Bush’s lead has fallen substantially from double digits to single digits, he has nevertheless held the lead over Gore in every Gallup survey from the first time Gallup tested his candidacy in May 1998. Bush has led Gore in 31 Gallup pre-election trial heat surveys since that time, including all 13 trial heats measured this year” (Gallup News Service).

While then-Governor Bush did end up winning the election in an electoral victory, the popular vote was still lost. This led to a dramatic Supreme Court case in which Bush was found as the victor of the presidential race. Although the polls accurately reflected the results of the presidential race, there is no doubt that the polls did not accurately reflect the actual results of the 2000 election.

2) Election of 2016

In one of the most dramatic elections of the century if not American history, Hilary Clinton was defeated by outside-contestant Donald J. Trump. It is unknown why this situation is. In a 2017 article published by the New York Times by Nate Cohn, it’s suggested that the polls were misleading due to the fact that the population sample didn’t accurately reflect the actual voter population.

“A postelection survey by Pew Research, and another by Global Strategy Group, a Democratic firm, re-contacted people who had taken their polls before the election. They found that undecided and minor-party voters broke for Mr. Trump by a considerable margin — far more than usual. Similarly, the exit polls found that late-deciding voters supported Mr. Trump by a considerable margin in several critical states. These three results imply that late movement boosted Mr. Trump by a modest margin, perhaps around two points” (Cohn).

In addition to not accounting for undecided voters, polls did not reflect the amount of last-minute changes there would be from voters in that base.

3) Election of 2020

Although Joe Biden was able to win both the electoral and popular vote by a secure amount of votes, polls did not reflect the actual results of the election either. Many polls forecasted over a 12 percentage point Biden victory while the results, at least in the electoral count, were in the razor-thin margins. A 2021 study by the Pews Research Center, “What 2020’s Election Poll Errors Tell Us About the Accuracy of Issue Polling” by Scott Keeter, Nick Hatley, Arnold Lau, and Courtney Kennedy questioned exactly that, misleading polls:

“The true picture of pre-election polling’s performance is more nuanced than depicted by some of the early broad-brush postmortems, but it is clear that Trump’s strength was not fully accounted for in many, if not most, polls. Election polling, however, is just one application of public opinion polling, though obviously a prominent one. Pollsters often point to successes in forecasting elections as a reason to trust polling as a whole. But what is the relevance of election polling’s problems in 2020 for the rest of what public opinion polling attempts to do? Given the errors in 2016 and 2020, how much should we trust polls that attempt to measure opinions on issues?” (Keeter, et al.).

II. Additional Marginal Elections

In the case of 1824, Andrew Jackson had won the election by the popular vote; he had however failed to achieve the majority of the electoral vote, leading the election process to be passed

into the House for voting. The house at the time was held by Henry Clay (then-Speaker of the House) and stern Adams-supporters which unsurprisingly voted in favor of Adams in the run-off election.

The 1876 election saw an incredibly close election with Hayes winning with just 1 electoral vote; he had however lost by over 200,000 votes in the popular vote, losing around 2.1% of the total popular vote count. This narrow election was only melded by Hayes making concessions to the Southern Democratic Party by ending the Reconstruction Eras effectively, ending Northern military control of the South and beginning the long-stretched Jim Crow era.

A more controversial election came in 1888 when incumbent President Grover Cleveland lost the electoral vote despite winning 2.2% more of the popular vote. Tariff policies had been key in the election and Cleveland's stance on reducing tariffs had led to the northern industrialists (specifically those centered around the Great Lakes) to vote for Harrison.

Although data on polling are too outdated to provide connections for this study, these elections showcase a similar pattern that raises the same question: had polls existed, would they have indicated radically different results from what the predicted information was?

III. Misleading Polls

So, why are polls in American politics so inaccurate? Why was it that the Hilary Clinton campaign predicted they'd win in a landslide victory but won the popular vote by significantly less and lost the electoral vote by a complete landslide? Perhaps the answer lies in the poles. In a 2020 article by Gloria Dickie published in the Scientific American, Dickie reflects and provides an explanation for the polling errors that occurred in the 2016 presidential election:

"The big polling errors in red states are the easiest to explain because there's a precedent: in states that are historically not very close for the presidency, the winning candidate usually overperforms. It's long been known turnout is lower in states that aren't competitive for the presidency because of our weird Electoral College mechanism. That effect—the winner's bonus—might be enhanced in very red states by the pandemic. If you're in a very red state, and you're a Democratic voter who knows your vote doesn't affect the outcome of the presidential race, you might be slightly less motivated to turn out during a pandemic" (Dickie).

This article suggests that although the polls correctly reflect the opinions of the American people, voters who turn out during elections don't necessarily reflect that.

"That's one kind of polling error that I don't think we need to be concerned about. But the error we probably should be concerned about is this 2.5-percentage-point error in close states. That error happened in swing states but also in Democratic-trending states. For people who watch politics closely, the expectation was that we had a couple of roads we could have gone down [on election night]. Some states count and report votes on election night, and other states take days to report. The polls beforehand pointed toward the possibility of North Carolina and Florida coming out for Biden. That would have effectively ended the presidential race right there. But the races were close enough that there was also the possibility that things would continue. In the end, that's what happened: we were watching more counting happen in Pennsylvania, Michigan, Wisconsin, Arizona and Nevada" (Dickie).

Ultimately, key states were flipped on election night because the polls simply were incorrect.

IV. Present-Day Issues with Early Calls

Why is this an issue in the first place? The issue lies with early calls. While the Dewey-Truman situation was one that was more humorous, there could be further issues with misleading polls. In "Early Calls of Election Results and Exit Polls: Pros, Cons, and Constitutional Considerations," by J. Ronald Milavsky, Al Swift, Burns W. Roper, Richard Slant, and Floyd Abrams in 1985, the authors stated the following regarding the relationship between close elections and

"The greater the margin between the winner and the loser, the sooner in the day it is possible to announce. Conversely, the smaller the margin between the winner and loser, the later in the day it is possible. When calls are made late, as they were in the presidential election of 1976, there is no problem. Problems may arise when calls are made before the polls close in many states. There is at least a possibility that the behavior of some voters, particularly those in western states, may be affected." (Milavsky).

In its most radical states, these early calls can lead to voters not even trying to vote as they may see their vote as being unimportant in the total result of the election.

V. Polarization

In his farewell address, President George Washington outlined the dangers of forming political parties and the destruction of political polarization the nation would have. Despite being aligned with the Federalists of his time, Washington remained moderate, having a cabinet that represented most spectrums in American politics. Unfortunately, immediately following the absence of Washington, the polarization of politics in America began. The Revolution of 1800 in which the Democratic-Republican Party captured the White House ultimately led to the demolition of the Federalist Party. The following decades would lead to a political fight between the Democratic-Republicans and Whigs arguing over foreign policy in the growing nation. While the post-Civil War Era saw a prolonged era of relative political stabilization with the Republican Party holding strong dominance from the 1860s to the 1930s, American politics slowly ran down the path of mutual destruction.

The modern Democrats and Republicans throughout the late 20th Century would disagree on virtually everything and the final decade of the 20th Century, marked by the Clinton Administration, would see a time of political unrest with the 104th Congress symbolizing the American government's instability. Even with the short-lived Republican dominance with the Regan, H.W. and Bush Administrations, there was never a prolonged period of complete dominance for either party with parties interchanging after almost every president (with the exception of the Roosevelt-Truman, Kennedy-Johnson, and Reagan-Bush administration shifts). In "America's Missing Moderates: Hiding in Plain Sight," Morris P. Fiorina offers an explanation as to why American politics have become so unstable.

Fiorina first suggests an increasing divide in the ideologies of the "political class, including convention delegates, donors and campaign activists." With such divide in the political establishments, the viewpoints of the politicians that the American people see on national television have become significantly polarized; this accounts for the substantive influence that has led to the polarization that the common people have gone through. Fewer moderates now enter public office, leading many congressional districts during elections to lack a moderate candidate that people can vote for. With this, election results have only gotten more polarized.

A secondary explanation has been that the American electorate no longer has a home party. Fiorina states that "the distributions of partisanship and ideology have not changed shape for a generation." However, what has changed is the sorting of political views. In the past, many Americans aligned with parties that often diverged from their own political views; since the

advancements of social media and the press, Americans tend to vote for the wrong party, leading many to choose one party and stay with it. This polarizes individuals as many end up walking deeper into the core of each ideology, a plausible explanation for the rise of far left and right groups.

The 2012 Election serves as an example of diverging politics according to Fiorina. Since “the American people do not give mandates” but instead turn to hiring “parties provisionally and grant them a probationary period to prove their worth,” many politicians find themselves establishing themselves on Capitol Hill or the White House only to have their pass revoked due to their poor history. Despite Bush starting his first year rather strong in terms of approval with the 9/11 terrorist attacks and the ensuing military response, the end of his term in 2008 saw a shift towards progressivism with the Financial Crisis of 2007-2008 completely tarnishing the face of the Republicans.

On the other hand, in “Polarized or Sorted? Just What’s Wrong With Our Politics, Anyways?” Alan I. Abramowitz provides a response to Fiorina’s explanations for the divergences in American politics. Instead of pointing to the differentiations in political groups, Abramowitz instead points to the rise in minority groups within the voting base. The article presents that “between 1992 and 2012, the nonwhite share of the presidential electorate more than doubled, from 13 percent to 28 percent,” leading to a “growing racial divide between the Democratic and Republican electoral coalitions.” This change has been bolstered by a change in societal norms: “the decline of the traditional family, the growing economic independence of women, and the rise of the women’s and gay rights movements” has led to an era of new ideologies. With different political beliefs on the table, parties have been quick to respond to them by adjusting their stances on many viewpoints in order to win a bigger electoral coalition. This has ultimately led to a strong polarization as the far wings of both spectrums have increased greatly: “white evangelical Christians now make up one of the largest and most loyal components of the Republican electoral coalition while secular voters overwhelmingly support Democrats.”

Additionally, instead of pointing to a sorting of voters to account for the changes in how people vote, Abramowitz points to the distribution of the voters on the ideology scale as also becoming more polarized. Unlike Fiorina who chose to blame the political class itself for being very polarized, Abramowitz points to the individual voters as a source of such drift. The article presents that “The distance between the average Democratic voter and the average Republican voter on the 7-point ideology scale more than doubled, from .9 units in 1972 to 2.2 units in 2008 with the average Republican voters going from a mean score of 4.7 in 1972 to a mean score of 5.4 in 2008 and the average Democratic voter going from a mean score of 3.7 in 1972 to a mean score of 3.2 in 2008.” This is complementary to the initial point that Abramowitz pointed out as the demographics have changed, making it only sensible that the viewpoints of the American people would also change. A shifted demographics in America with the rise of immigration in the late 20th century would also make plausible sense.

Viewing these political and demographic changes from a historical lens, the dangers of such shifts are evidently seen. Even during the landslide election victories by Presidents Franklin D. Roosevelt, Richard Nixon, and Ronald Reagan, there has never been such polarization in how the American people view the two political parties. The only instance in which something similar has ever occurred in American history are the decades leading up to the Civil War. The established parties saw great upheavals with the Democratic-Republicans, the Whigs, and the Know-Nothings losing their stances with the topic of abolishment becoming the most prevalent topic. At the center of the historical connection is the fact that the Republican and Democratic parties have both lost their initial identities. From the party of Abraham Lincoln to the party of Donald Trump, the Republican party has gotten increasingly conservative while originating from white supremacy in the

South to eventually Barack Obama, the Democratic Party has become one of the most progressive parties in the history of America.

Polarization has occurred on numerous instances throughout political history, whether it be strictly American or international. However, given the rapid rates in which parties have become shifted to the far wings of their set spectrums, it is highly likely that in the recent future there will be a meltdown of ideologies, party alignments, and political culture. While it's clear that America will get back on its feet even with polarization, chaos seems to be inevitable. And at the center of the house stands a hidden devil: polls.

VI. Morality of Polling

In viewing the morality of polling overall, it is essential that pollsters do not release misleading data as it can hinder an individual's motivation to vote as well as entire campaign teams on where to lead their last few days.

Additionally, from the perspective of the individuals who decide to vote, it is key that discouragement does not come from these votes as every vote truly matters - voting should not be dependent on the gravitation of the extent of their political power.

Part III: Search Engines (SEOs)

I. Introduction

Data Availability also plays a major role in search engines (SEOs) as well. As of January of 2022, "Google processes over 3.5 billion searches every day, or, 1.2 trillion searches per year" (Chaffey). In addressing the rise of data available for daily use, it is rather coherent to predict that this will lead to issues in decision-making. In "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips" by Betsy Sparrow, Jenny Liu, and Daniel M. Wegner, the experiments address this issue:

"The advent of the Internet, with sophisticated algorithmic search engines, has made accessing information as easy as lifting a finger. No longer do we have to make costly efforts to find the things we want. We can "Google" the old classmate, find articles online, or look up the actor who was on the tip of our tongue. The results of four studies suggest that when faced with difficult questions, people are primed to think about computers and that when people expect to have future access to information, they have lower rates of recall of the information itself and enhanced recall instead for where to access it. The Internet has become a primary form of external or transactive memory, where information is stored collectively outside ourselves" (Sparrow et al.).

The abstract suggests that information that was usually stored within ourselves are now "stored collectively outside ourselves." What are the implications of this issue though? If data is being stored outside ourselves, how can we even access this information on a daily basis and what effects can this have on our mind process? The study continues on:

"These studies suggest that people share information easily because they rapidly think of computers when they find they need knowledge (Expt. 1). The social form of information storage is also reflected in the findings that people forget items they think will be available externally, and remember items they think will not be available (Expts. 2 and 3). And transactive memory is also evident when people seem better able to remember which computer folder an item has been stored in than the identity of the item itself (Expt. 4). These results suggest that processes of human memory are adapting to the advent of new computing and communication technology. Just as we learn through transactive memory who knows what in our families and offices,

we are learning what the computer “knows” and when we should attend to where we have stored information in our computer-based memories. We are becoming symbiotic with our computer tools (8), growing into interconnected systems that remember less by knowing information than by knowing where the information can be found. This gives us the advantage of access to a vast range of information—although the disadvantages of being constantly “wired” are still being debated (9). It may be no more that nostalgia at this point, however, to wish we were less dependent on our gadgets. We have become dependent on them to the same degree we are dependent on all the knowledge we gain from our friends and coworkers—and lose if they are out of touch. The experience of losing our Internet connection becomes more and more like losing a friend. We must remain plugged in to know what Google knows” (Sparrow et al.).

If we are being plugged into what Google knows, is that for the better? Additionally, if much more evidence is within the reach of our fingertips, is that really beneficial? Will having access to limitless amounts of information and evidence to base off our decisions truly be better in formulating better social decisions? Conclusively though, it is evident that individuals’ reliance on data has significantly changed the approach individuals take on information processing and game theory.

Part IV: Nutrition Labels

I. A Study: Consumption of Calcium and Knowledge About Calcium Sources and Nutrition Labels Among Lower Secondary Students in Thailand

1) Context

Raksaworn Jaissard, Tipaporn Kanjanarach, Sutin Chanaboon, and Borey Ban looked into how much calcium Thai secondary school children were consuming. Their findings were published online on September 14th, 2021 in the National Library of Medicine. This investigation was divided into three phases:

- 1) The initial investigation focused on locating high-calcium goods in the consumer's neighborhood market.
- 2) In the second study, it was determined which high-calcium products lower secondary Thai pupils were aware of.
- 3) Calculating how much calcium the students had actually consumed during the previous 24 hours was the focus of the third investigation.

In light of this, the research's straightforward goal was to determine how much calcium lower secondary kids have access to, are aware of, and actually eat. All of the students were in the age range of 11 to 14 and in grades 7 to 9. Approximately 309 students were invited with 168 students in study 1 and the other 209 in study 2.

2) Findings

It was found that within the native market, there have been around ninety four total high-calcium products. However, only around forty nine were known by the scholars as “known” or consumed. “The median quantity of calcium consumed within the previous twenty four hours was 410 mg with solely around thirty one.1% students (65/209) intense what the govt began as daily recommendations.

As expected, the most supply of calcium was through milk, furthermore as malt drink, minnow, canned fish, calcium fortified soy milk, and Chinese kale. “The overall average data score was half dozen.5±2.4 out of a maximum of seventeen. the common data scores for top calcium

food sources and for decoding the calcium info on a sample nutrition label were 4.6 ± 1.8 (out of 12), and 1.9 ± 1.2 (out of 5), respectively” (Jaisaard et al.).

3) Study Relevance

With such an occasional proportion of scholars having the ability to spot high-calcium products and really intense them, it's evident that young kids don't seem to be obtaining access to the nutrients that they have. With calcium being the foremost essential mineral, we are able to draw an identical conclusion that different minerals and nutrients key to childhood growth are in deficit.

II. Morality

What do individuals do with datasets they are given? In any normal debate over the data set of something like nutrition, it should be expected that people are always striving *for* data. However, we see in the Thai student example that individuals, despite having some data necessary to make a good decision with nutrition and health, choose an option that's less viable.

Part VI: Conclusion

In specifically politics, it is rather clear that data can drive individuals to not vote in a key election. As a result, the reality can often differ from the conclusion that polls originally forecasted as we saw with previous elections. In the case of SEOs, having data and information generally seems to be a boost to decision-making with individuals having access to more evidence and intelligence to go off of. However, with nutrition, there were evident playoffs between having data and information; it was evident that individuals don't always make the “morally” correct decisions despite having the data.

In any case, it's rather clear that information and datasets play key roles in an individual's decision making process and it is important for both the data provider as well as the receiver to consider discrepancies in the data.

Works Cited

- Chaffey, Dave. "Search Engine Marketing Statistics 2022." *Smart Insights*, 26 Jan. 2022, <https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/#:~:text=As%20of%2026%20January%202022,1.2%20trillion%20searches%20per%20year.&text=Moreover%2C%20mobile%20shopping%20now%20has,300%2B%20page%20report%20on%20Data%20portal.>
- Cohn, Nate. "A 2016 Review: Why Key State Polls Were Wrong about Trump." *The New York Times*, The New York Times, 31 May 2017, <https://www.nytimes.com/2017/05/31/upshot/a-2016-review-why-key-state-polls-were-wrong-about-trump.html>.
- Dickie, Gloria. "Why Polls Were Mostly Wrong." *Scientific American*, Scientific American, 13 Nov. 2020, <https://www.scientificamerican.com/article/why-polls-were-mostly-wrong/>.
- Gallup, Service. "The 2000 Presidential Election -- a Mid-Year Gallup Report." *Gallup.com*, Gallup, 18 Oct. 2014, <https://news.gallup.com/poll/9898/2000-presidential-election-midyear-gallup-report.aspx>.
- Keeter, Scott, et al. "What 2020's Election Poll Errors Tell Us about the Accuracy of Issue Polling." *Pew Research Center Methods*, Pew Research Center, 8 Apr. 2021, <https://www.pewresearch.org/methods/2021/03/02/what-2020s-election-poll-errors-tell-us-about-the-accuracy-of-issue-polling/>.
- Lemelin, Bernard. "The U.S. Presidential Election of 1948: The Causes of Truman's 'Astonishing' Victory." *Cairn.info*, Belin, 2001, <https://www.cairn.info/revue-francaise-d-etudes-americaines-2001-1-page-38.htm>.
- Milavsky, et al. "Early Calls of Election Results and Exit Polls: Pros, Cons, and ... - JSTOR." *JSTOR*, 1985, <https://www.jstor.org/stable/2748928>.
- Milavsky, J. Ronald. "Early Calls of Election Results and Exit Polls: Pros, Cons, and Constitutional Considerations: Introduction." *OUP Academic*, Oxford University Press, 1 Jan. 1985, <https://academic.oup.com/poq/article-abstract/49/1/1/1937341?redirectedFrom=PDF&login=false>.
- Sparrow, et al. "Google Effects on Memory: Cognitive Consequences of ... - Science | AAAS." *Science*, 2011, <https://www.science.org/doi/10.1126/science.1207745>.
- Sparrow, et al. "Dirección De Personal." *Dirección De Personal*, 2011, <http://www.personal.umich.mx/>.