Spring 5-8-2017

The Hidden Trump Model: Modeling Social Desirability Bias through Agent Based Models

Hannah Zontine

THE HIDDEN TRUMP MODEL: MODELING SOCIAL DESIRABILITY BIAS THROUGH AGENT BASED MODELS

An honors paper submitted to the Department of Computer Science

of the University of Mary Washington

in partial fulfillment of the requirements for Departmental Honors

Hannah Zontine

May 2017

By signing your name below, you affirm that this work is the complete and final version of your paper submitted in partial fulfillment of a degree from the University of Mary Washington. You affirm the University of Mary Washington honor pledge: "I hereby declare upon my word of honor that I have neither given nor received unauthorized help on this work."

Hannah Zontine
(digital signature)

05/08/17

The Hidden Trump Model

By Hannah Zontine

Section 1: Background

The majority of poll results published before the Presidential Election predicted Donald Trump's defeat 'by a landslide' (Pomarico 2016). On Election Day, in spite of this, he secured more than enough electoral votes to win him the presidency. What caused poll results to differ so significantly from election results? The sole purpose of polls is to represent the voting intentions of the population. Polling every eligible voter would require unfathomable resources; as a result, all polls utilize only a sample of the population. Unless pollsters fail to accurately sample from the voting population, their results should be similar to the votes each candidate receives. This issue came to light when Franklin D. Roosevelt beat Alfred Landon in the 1936 Presidential election. Pollsters realized that by conducting the poll explicitly over the phone, they were choosing to sample from voters who could afford a house phone; they neglected to poll from an accurate sample of the voting population. Since that election, pollsters have aimed to not repeat that mistake.

So what if there is a different effect at work that caused poll results to significantly differ from election results? What if people told pollsters their intentions were to vote against Trump, but, on election day, voted for him?

Section 1.1 The Bradley Effect

In 1982, the election for California's next governor was between Tom Bradley and George Deukmejian. Bradley, long-time mayor of Los Angeles, was African-American. Deukmejian was Caucasian. Bradley led the opinion polls throughout the election; naturally, everyone assumed he would win. However, once the votes were counted, Deukmejian was revealed the winner.

After Bradley's shocking loss, California residents wondered why the poll results differed from the election results. One theory persisted: people must have lied to pollsters. The "Bradley Effect (BE)," as it is now commonly known, is the "notion that voters overstate their support for a black candidate to pollsters for fear of being perceived as racist" (D'Aprile 2008).

In the early 1990s, Harvard political scientist Dan Hopkins studied the Bradley Effect by conducting a large-scale empirical study on it over many different elections. He claimed that the BE had a significant effect on the polls. He observed its negative effect on black candidates and reported a median gap (i.e. difference between polls and results) of 3.1 percentage points. However, after 1996, he found no evidence of the BE influencing any elections and reported a median gap of -0.3 percentage points.

Some speculated the Bradley Effect might have been at work against Barack Obama in the 2008 Presidential Election. Not only was he an African American running against a Caucasian, but he was the candidate leading in opinion polls. Nevertheless, he ended up winning both presidential elections he ran in.

Empirical evidence of the Bradley Effect rests on individual cases. While typically the BE is found to result in the African-American candidate losing, it was observed to have the opposite effect on the NYC 1989 mayor race. David Dinkins, an African-American candidate, was leading by a few percentage points against Ed Koch, the Caucasian incumbent, in the weeks leading up to the election. He ended up winning the election by over 8 percentage points despite the opinion polls suggesting a close race.

There is mixed evidence to support the influence the Bradley Effect can have over any type of election. Gary Langer, the director of polling for ABC news, believes the BE to be "a theory in search of data" (Wang 2008).

The Bradley Effect is explicitly about race. However, more generally, polled voters may be dishonest to pollsters for a variety of other reasons having to do with the perceived social stigma of supporting a certain candidate. In research, we can generalize the BE to all situations in which the bias exits.

Section 1.2 Social Desirability Bias

Social psychologists have long been interested in understanding why people lie to their friends and family about their true beliefs. They speculate that some people would prefer to lie about their opinions than face judgment or rejection from a close associate.

Social desirability bias is a term for the idea that the "basic human tendency to present oneself in the best possible light can significantly distort the information gained from self-reports" (Maccoby and Maccoby 1954). People are often unwilling to truthfully report their beliefs on "sensitive topics for ego-defensive or impression management reasons." Maccoby and Maccoby determined in 1954 that data collected from self-reports is "systematically biased toward the respondent's perceptions of what is socially acceptable." This phenomenon has been found to occur in "virtually all types of self-report measures and across nearly all social sciences literatures" (Maccoby and Maccoby 1954).

Section 1.3 The 2016 Presidential Election

The election for the 45th President of the United States took place on November 7, 2016. Over 138 million people visited a polling booth that day to submit a ballot (Levine 2016). While there were many candidates on the ballot, by far most voted for Hillary Clinton and Donald Trump. Clinton, the Democratic nominee, is a former Secretary of State, Senator, and First Lady.

Trump, the Republican nominee, is an American businessman, television personality, and politician.

Both candidates faced potential charges during their campaign that could have prevented them from being elected. For instance, Clinton was being investigated for emailing classified information through an unsecure, private server while she was the Secretary of State; however, the FBI dropped the investigation, stating "no charges are appropriate in this case" (Comey 2016). Trump faced multiple allegations of sexual assault, but was never formally charged.

While many people supported their candidate by publicly advocating for them, some people felt pressure not to. This was especially true of Donald Trump. Many of his supporters refused to express their voting intentions, because they either felt it would turn away their Democratic friends and colleagues or they would be "seen as culturally insensitive" (Simmons 2017; also see Shepard 2016, Enns & Schuldt 2016).

During the campaign, the majority of opinion polls predicted Clinton would win the election. Nate Silver, a political analyst who accurately predicted how 49 states would vote in the 2008 election, announced in his final election forecast that Clinton was a "71% favorite" in polls. Alan Abramowitz, a political scientist and author, wrote that "Clinton is heading for a decisive victory over Donald Trump based on national and swing state polls" (Abramowitz 2016). The HuffPost Presidential Forecast Model determined Clinton had a "98.2% chance of winning," while Trump had "no path to an Electoral College victory" (Jackson 2016). The model projected Clinton garnering 323 electoral votes and a shift in the Senate to a Democratic majority.

Trump's campaign manager, Kellyanne Conway, was among the first to publicly speculate that opinion polls weren't accurately reflecting the amount of support Trump had. In an interview with UK "Channel 4" news, she stated that "it's become socially desirable -- especially if you're a college-educated person in the United States of America -- to say that you're against Donald Trump" (Wright 2016). According to the campaign's research, Trump "performs consistently better in online polling" (Wright 2016).

The only major poll to predict Trump's win was the Daybreak Poll conducted by the LA Times. Based on an Internet probability survey, participants were sampled from an ongoing UAS (Understanding America Study) panel of 6,000 randomly selected U.S. residents. Selected residents who lacked Internet access were provided with some. Every day pollsters asked 1/7th of their participants three questions: 1) How likely are you to vote? 2) Which candidate do you think would win if the election was today? 3) Which candidate would you vote for if the election was today? The pollsters discovered that while more of the participants thought Clinton would win, most actually planned on voting for Trump. Their approach differed from other polls in that they adjusted their data so it would represent the diversity of the population.

There is good evidence to support the idea that for many Trump supporters their voting intentions differed from the ones they communicated to pollsters. Surely, this is not the first example of this phenomenon. The goal of my study was to model this more complex

decision-making process and determine the extent to which it accurately predicts real-world political outcomes.

Section 2: Opinion Dynamics Models

Simulations that aim to reproduce the phenomenon of individual agents forming opinions over time via mutual influence are known as Opinion Dynamics models. They are computational realizations of the structural balance theory from Cartwright and Harary in 1956, where individuals' opinions are influenced by "those with whom they share affective social ties" (Moore et al. 2015). These models draw upon concepts underlying Asch's famous experiments on opinion formation and conformity under social pressure (Asch 1955), and French's work on social power in networks (French 1956).

Section 2.1 Binary Voter Model

The Binary Voter Model (BVM) laid the initial foundation from which many other Opinion Dynamics models have been constructed (Clifford and Sudbury 1973; Holley and Liggett 1975). The BVM represents an artificial society of socially connected people. Individuals (in the terminology of Social Science Simulation, **agents**) hold a single opinion, which can be one of two different values. In a political sense, an agent would be viewed as a Liberal or Conservative.

The BVM simulates encounters between agents, some of whom are socially related to others. Periodically, a randomly chosen agent adopts the opinion of a neighboring agent if the two opinions differ. Over time, this always results in uniformity of opinion (Aldous and Fill 2002, ch. 14). However, in the real world, we observe that society doesn't always converge to uniformity. While the BVM is well-founded, the model is too simplistic to accurately reflect the dynamics of a real society's opinions. As a result, researchers in the field of Opinion Dynamics have implemented a plethora of variations on the model.

Originally, the social relations were depicted as a regular lattice, which means that agents were related to other agents in a predictable, uniform way. Subsequent variations to the model have depicted agents dispersed on an arbitrary graph.

Other extensions were discovered to keep a society from converging on a single opinion. In one, Yildiz altered the BVM so that some agents in the graph never changed their opinion (Yildiz, Acemoglu, Ozdaglar, Saberi, Scaglione 2011). In the presence of only a few so-called **stubborn agents**, this variation was determined to always resulted in a society of polarized opinions.

A different extension of the BVM replaces the agent's binary opinion value with a continuous one. If the difference of opinion between two agents is greater than a fixed threshold

value, they won't interact. Refused discussion is meant to represent a "lack of understanding, conflicting interest, or social pressure" (Weisbuch, Deffuant, Amblard, and Nadal 2001). If the threshold value is set low enough, two separate clusters of opinions will form. Agents will never interact with agents from the other cluster, and consensus of opinion will not be achieved.

Section 3: The "Hidden Trump Model"

In this thesis, I present the **Hidden Trump Model** (HTM), a novel extension to the Binary Voter Model. Every agent is assigned *two* binary opinion values, called **expressed** and **latent**. Each of an agent's opinions can take one of two possible values. For example, one interpretation is that an agent is either 'pro-Trump' or 'anti-Trump'.

An agent's expressed opinion is the one he communicates that he believes in face-to-face social situations, while an agent's latent opinion is the one that he truly believes. **True believers** are agents with identical expressed and latent opinions. Other agents, however, whose expressed and latent opinions differ, do not truly agree with the opinion they tell others they agree with.

Agents interact in two different environments: online anonymously with strangers, and face-to-face with acquaintances.

When an agent interacts anonymously, she will disclose her honest opinion without fear of judgment. Not all online interactions are anonymous, of course, but there is evidence to suggest that communicating online may lead to more open and "lively discussions as they plausibly allow participants to be anonymous and alleviate the hesitancy of those members holding the minority viewpoints to speak out" (Ho and McLeod 2008).

When an agent interacts face-to-face, on the other hand, his true political affiliation may be hidden behind the facade of his expressed opinion. Studies have shown that fear of isolation and future opinion congruence are significantly related to willingness to express one's own opinion (Ho and McLeod 2008). This model assumes that an agent may feign agreement with an associate if she perceives their opinions disagree.

Face-to-face encounters, in addition to possibly modifying what people may express, may also cause them to change their latent opinion. Thus, during an interaction, if two agents' expressed opinions already match, one of the two may "internalize" that opinion and become a true believer. Note that this aspect of the model causes the two variables to be intertwined -- without it, the model would reduce to two (predictably behaving) independent BVM processes.

These two styles of interaction represent idealized poles on a spectrum from fully anonymous to directly face-to-face. Not all real-life interactions are at one of these extremes, of course. All models are simplifications of reality, however, and the HTM approximates the continuum of interaction types with these two representative processes. In particular, I am modeling "online" interactions as "fully anonymous," which of course is a generalization of real-life communications.

What will result when multiple agents behave according to these two processes is non-obvious, and the focus of my thesis.

Section 4: The Hidden Trump Model in its Extremes

The influence of the two processes can be characterized by three probabilities. The impact of the online process is controlled by the **update probability** (*i.e.* the probability an agent will change his/her latent opinion to mimic the encountered agent's latent opinion, if they don't already match). The effect of the face-to-face process is determined by two probabilities. The first is the **peer pressure probability** (*i.e.* the probability an agent will change his/her expressed opinion to imitate the other agent's expressed opinion, if they disagree). Note that this is asymmetric: the "anti-Trump" opinion is the socially acceptable one, and peer pressure always exerts itself in that direction. The second is the **internalize probability** (*i.e.* the probability an agent will change his/her latent opinion to match his/her expressed opinion). This probability only comes into play when the two agents agree externally. These three key parameters can each be adjusted in the range 0 to 1.

It is interesting to consider what behavior the model will exhibit when each of these parameters is set to an extreme value -- 0 or 1. Obviously, if all three probabilities are set to 0, the agents' opinions never change (see table, case 000).

Process	Online	Face-to-Face	Face-to-Face	Effect on the Graph
Case	*Update Probability*	*Peer Pressure Probability*	*Internalize Probability*	
000	0	0	0	All interactions are ineffective Static latent and expressed opinions
001	0	0	1	Two ineffective processes Static expressed opinions, which latent eventually matches
010	0	1	0	Ineffective online process Static latent opinions Single BVM (on expressed attribute) with graph neighbors

011	0	1	1	Ineffective online process Expressed and latent opinions reach uniformity
100	1	0	0	Ineffective face-to-face process Static expressed opinions Single BVM (on latent attribute) with mean field
101	1	0	1	Society based solely on online social influence Effective online process Static expressed opinions Latent opinions polarize
110	1	1	0	Two independent effective processes Two single BVMs (on expressed and latent)
111	1	1	1	Full influence of HTE Two dependent effective processes

When only the internalize probability is set to 1 (see table, case 001), both online and face-to-face encounters are unsuccessful. If an agent meets another with whom he externally disagrees, nothing will happen. However, as soon as he interacts with an agent with whom he shares an expressed opinion, he will permanently change his latent opinion to match his expressed opinion. As the simulation runs, the expressed opinions of agents on the graph are never influenced and the latent opinions of agents update once to reflect the agent's expressed opinion.

When only the peer pressure probability is turned on (see table, case 010), the face-to-face process functions as a single Binary Voter Model on the expressed attribute with graph neighbors, which will eventually reach uniformity of expressed opinion. Since the online process is always unsuccessful, latent opinions remain static throughout the simulation.

When the face-to-face process is turned on (i.e. the peer pressure probability and internalize probability are set to 1), but the online process is turned off (i.e. update probability is set to 0), the graph will soon reach uniformity of the 'anti-Trump' opinion (see table, case 011). Expressed opinions of agents in the graph will steadily reach uniformity with latent opinions of agents following soon after. Initially, the majority of latent opinions may progress toward 'pro-Trump', but as soon as all the expressed 'pro-Trump' opinions die out, all the external 'anti-Trump' opinions will become internalized by agents.

When only update probability is set to 1 (see table, case 100), the online process functions as a single Binary Voter Model on the latent attribute with Mean Field. Since the face-to-face process is always unsuccessful, expressed opinions stay static and there is no influence of the HTE. Agents only interact online sharing their latent opinions with others. The latent opinions of the agents on the graph will eventually reach uniformity of opinion.

When the internalization probability and update probability are both turned on (see table, case 101), agents on the graph are only successfully influenced by other agents that they interact with anonymously through the online process. Without peer pressure influencing the face-to-face process, expressed opinions remain half 'pro-Trump' and half 'anti-Trump' throughout the simulation. Whenever two agents with the same expressed opinion interact face-to-face, the victim agent will internalize that shared opinion. Without the influence of the internalize probability, the online process would function as a single BVM and reach uniformity of latent opinion. However, because the internalize probability is set to 1, agents in the graph never converge on a single latent opinion. This combination of the probabilities depicts how a society, completely reliant on online communication, will result in agents with polarized latent opinions. A society polarized by opinion can never reach uniformity of opinion.

When only the internalize probability is turned off (see table, case 110), the two processes operate independently. The face-to-face process functions as a single Binary Voter Model on the expressed attribute and influencers are sampled from an agent's graph neighbors. Agents whose expressed opinions are 'pro-Trump' always conform when they encounter agents who hold 'anti-Trump' expressed opinion. As a result, expressed opinions quickly reach uniformity. The online process functions as a single Binary Voter Model on the latent attribute and influencers are sampled from the mean field. Latent opinions reach uniformity at a slower rate than expressed opinions.

When all three probabilities are turned on (see table, case 111), the face-to-face and online processes are dependent on one another. The desire to express the socially acceptable opinion has complete influence over the expressed opinions of the agents in the graph. Those expressed opinions will be internalized if the agent agrees externally with another agent during a face-to-face interaction. As a result, all expressed opinions become 'anti-Trump' and hidden opinions eventually follow. Not only is there an ongoing pressure from the internalize probability make latent opinions reflect expressed opinions, but an unbiased BVM process is influencing latent opinions as well.

Section 5: Results

There are four parameters to the Hidden Trump Model: graph density, update probability, peer pressure probability, and internalize probability. The influence of these parameters can be measured with the difference between an opinion poll (where the society of agents is the sample

population) and the results of the election (where the society is the voting population). If a pollster asked every agent for their opinion, they would share their expressed opinion. However, agents in private voting booths would vote according to their true (latent) opinions. Poll bias is a term that represents the difference between the percentage points a poll reports for a candidate and the amount of votes that candidate actually receives in the election. I compute this value using the number of 'pro-Trump' expressed and latent opinions. The more agents who express 'anti-Trump' but believe 'pro-Trump' the higher a society's poll bias will be.

For each of the four parameters, I measure its influence on the graph, while keeping the other parameters constant. I ran 50 simulations for each and recorded the maximum poll bias.

Graph Density

The density of a graph is defined as the number of connections between agents divided by the number of potential connections between agents. Initializing a random graph requires a probability of connection between agents. As the probability increases, agents have a larger social circle. Of the two processes, only the Face-to-Face process is influenced by this probability, because agents only interact with agents they are connected to.

The other three probabilities are held constant at 0.5.

Probability of Connection between Agents
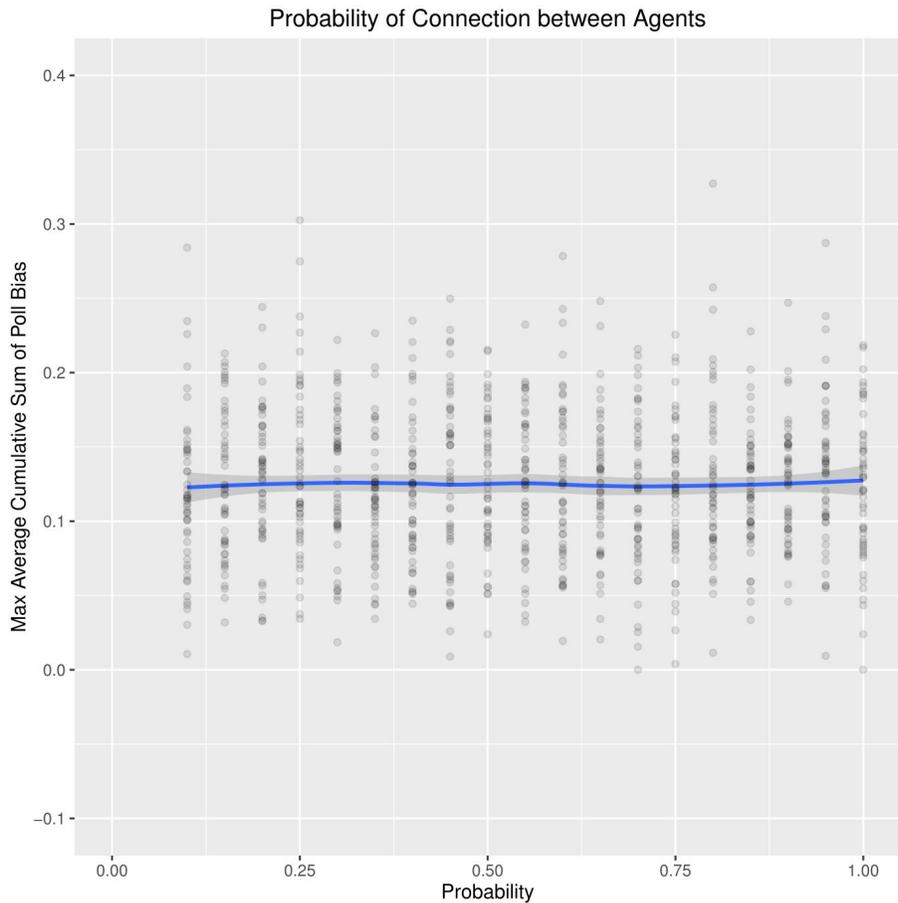
This model suggests that poll bias is not affected by the size of an agent's social circle.

Peer Pressure Probability

The peer pressure probability is an essential component of the Face-to-Face process. As the probability increases, agents are more likely to succumb to peer pressure and modify their expressed opinion.

Internalize probability and update probability are held constant at 0.5, while the density of the graph is held at 0.25.
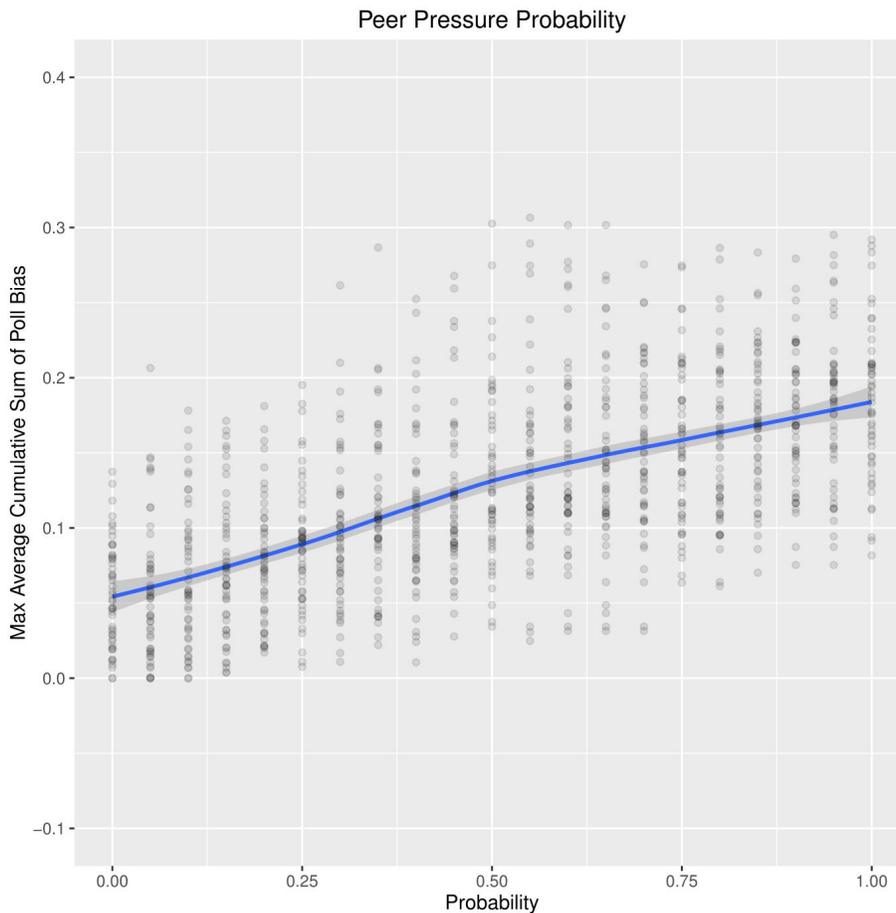

This model suggests that even the influence of a mild peer pressure environment can cause a significant poll bias to occur.

Internalize Probability

The internalize probability is the other essential component to the Face-to-Face process. As the probability increases, agents are more likely to change their latent opinion to match their expressed opinion.

Peer pressure probability and update probability are held constant at 0.5, while the density of the graph is held at 0.25.
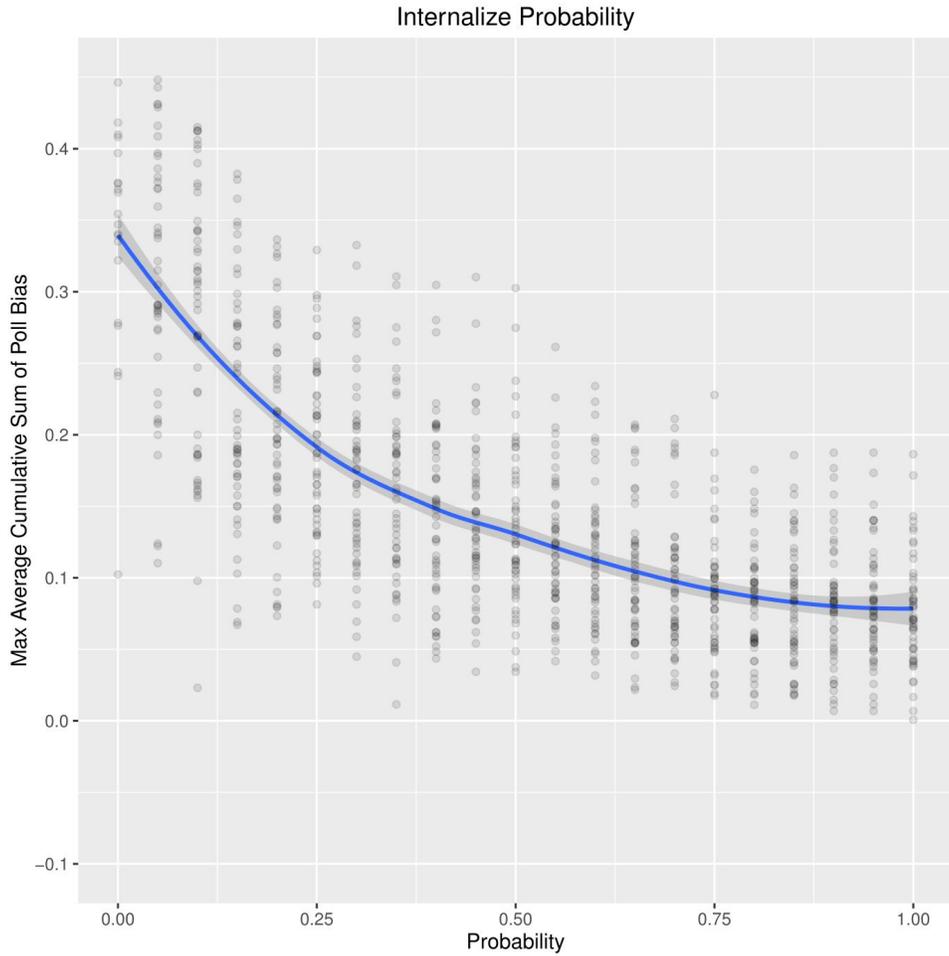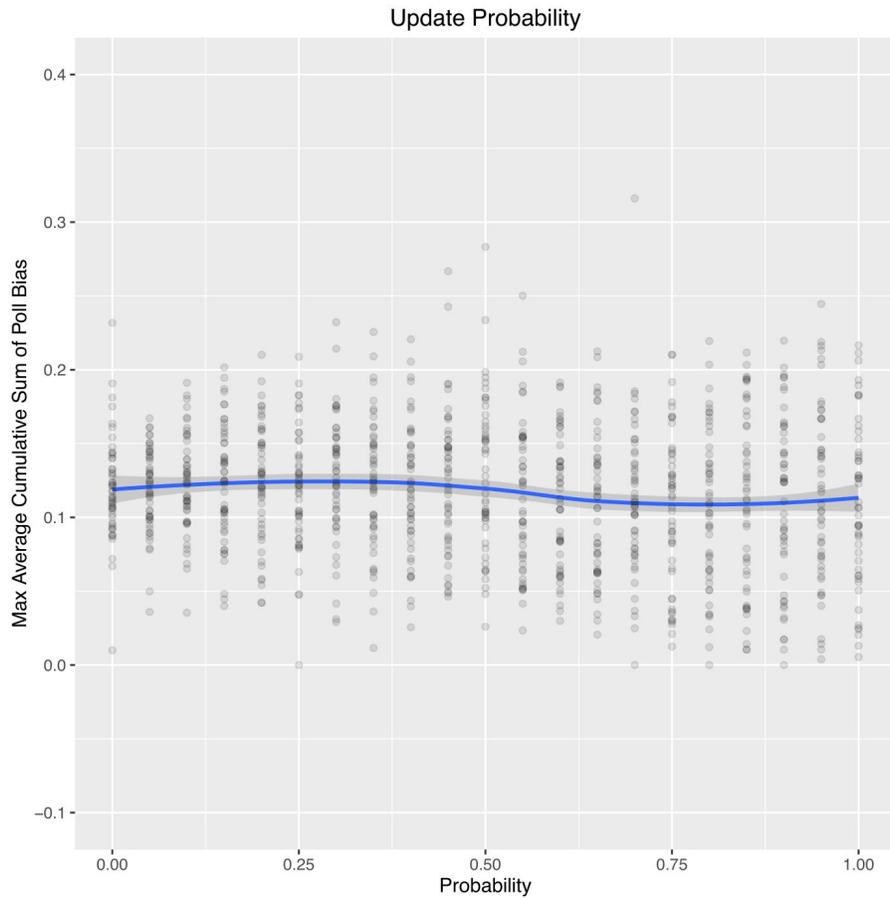

This model suggests that as this probability and poll bias are negatively correlated: as internalize probability increases, poll bias decreases. Agents are more aggressively copying their expressed opinion to become their real opinion; in short, agents are being more honest. This model shows that a high poll bias can occur if agents, who support the candidate with stigma, experience stronger internal resistance toward supporting the other candidate even if they're saying they do.

Update Probability

The update probability is sole component to the Online process. As the probability increases, agents are more likely to change their latent opinion to match the other's agent's opinion during anonymous interactions.

Peer pressure probability and internalize probability are held constant at 0.5, while the density of the graph is held at 0.25.

This model suggests that poll bias is not affected by this probability.

Section 6: Conclusion

In conclusion, this model suggests that poll bias is not actually affected by people's interaction frequencies. Getting people to interact more online or expand their social circles won't prevent inaccurate polls. Instead, the solution to poll bias is people having less resistance toward being truthful; in an environment free of peer pressure, they will.

Works Cited

Abramowitz, Alan I. "History Shows Hillary Clinton Will Win." *NY Daily News*. N.p., 01 Nov. 2016.
        Web. 06 Jan. 2017.
        <http://www.nydailynews.com/news/politics/history-shows-hillary-clinton-win-article-1.2854411
        >.

Aldous, David and Fill, James Allen 2002. "Reversible Markov Chains and Random Walks on Graphs".
        Unfinished monograph, recompiled 2014, available at
        http://www.stat.berkeley.edu/~aldous/RWG/book. html.

Asch, S.E. (1955). Opinions and Social Pressure. *Scientific American, 193*, 31–35.
        [doi:10.1038/scientificamerican1155-31]

Clifford, P., and A. Sudbury. 1973. "A Model for Spatial Conflict". *Biometrika* 60 (3): 581–588.

D'Aprile, Shane. "Shattering the Bradley Effect?" *Campaigns & Elections*. 3 Nov. 2008. Web.
        <https://www.campaignsandelections.com/campaign-insider/shattering-the-bradley-effect>.

French, J.R., & R.P. John. (1956). A Formal Theory of Social Power. *Psychological Review, 63*,
        181–194. [doi:10.1037/h0046123]

Jackson, Natalie. "HuffPost Forecasts Hillary Clinton Will Win With 323 Electoral Votes." *The
        Huffington Post*. N.p., 7 Nov. 2016. Web. 6 Jan. 2017.
        <http://www.huffingtonpost.com/entry/polls-hillary-clinton-win_us_5821074ce4b0e80b02cc2a94
        >

Jung, Jiin, and Aaron Bramson. *A Recipe for Social Change: Indirect Minority Influence and Cognitive
        Rebalancing* (2016): n. pag. Web.
        <http://computationalsocialscience.org/wp-content/uploads/2016/11/CSSSA_2016_paper_5-1.pdf
        >.

Holley, and Liggett. 1975. "Ergodic theorems for weakly interacting systems and the voter model".

Holmes, Steven. "Pollsters Debate 'Bradley Effect'." *The Washington Post*. WP Company, 12 Oct. 2008.
        Web.
        <http://www.washingtonpost.com/wp-dyn/content/article/2008/10/11/AR2008101102136.html>.

Levine, Daniel S. "Over 90 Million Eligible Voters Didn't Vote in the 2016 Presidential Election."
        Heavy.com. N.p., 10 Jan. 2017. Web.
        <http://heavy.com/news/2016/11/eligible-voter-turnout-for-2016-data-hillary-clinton-donald-trum
        p-republican-democrat-popular-vote-registered-results/>.

Maccoby, E. E., & Maccoby, N. (1954). The interview: A tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology: Vol. 1. Theory and method* (pp. 449-487). Reading, MA: Addison-Wesley.

Moore, Thomas, Finley, Patrick, Brodsky, Nancy, Brown, Theresa, Apelberg, Benjamin, Ambrose, Bridget and Glass, Robert (2015) 'Modeling Education and Advertising with Opinion Dynamics' Journal of Artificial Societies and Social Simulation 18 (2) 7 <http://jasss.soc.surrey.ac.uk/18/2/7.html>. doi: 10.18564/jasss.2680

Pomarico, Nicole. "Nate Silver's Election Prediction Says Hillary Clinton Will Win By A Landslide." *Bustle*. Bustle, 19 Oct. 2016. Web. <https://www.bustle.com/articles/190473-nate-silvers-election-prediction-says-hillary-clinton-will-win-by-a-landslide>.

Shepard, Steven. "GOP Insiders: Polls Don't Capture Secret Trump Vote." *POLITICO*. N.p., 28 Oct. 2016. Web. <http://www.politico.com/story/2016/10/donald-trump-shy-voters-polls-gop-insiders-230411>.

Silver, Nate. "Final Forecast." *Twitter*. Twitter, 08 Nov. 2016. Web. <https://twitter.com/NateSilver538/status/796000742066753537>.

Simmons, Christine. "For Many Big Law Trump Donors, 'Stigma' Kept Support Below the Radar." *Law.com*. N.p., 16 Jan. 2017. Web. <http://www.law.com/sites/almstaff/2017/01/16/for-many-big-law-trump-donors-stigma kept-support-below-the-radar/?slreturn=20170316221250>.

Wang, Sam. "The Disappearing Bradley Effect." *Princeton University*. The Trustees of Princeton University, 28 Sept. 2008. Web. <http://election.princeton.edu/2008/09/27/the-disappearing-bradley-effect/>.

Wright, David. "Campaign Manager Points to 'hidden Trump Vote'." *CNN*. Cable News Network, 24 Aug. 2016. Web. <http://www.cnn.com/2016/08/24/politics/donald-trump-undercover-voters-kellyanne-conway/>.