

Politician Ratings Website and Experiment

Project Overview

This project is based on Ben Buras' suggestion: that he could help create "a truth ratings website that would attempt to set truth ratings for politicians."

Here's a summary of how this project could work:

1. We create a tiny site that can serve as a prototype of what a full blown Truth Ratings site would look like.
2. We get lots of people's opinions on the prototype site. We use that to improve it to the point where they say "Yes, this would cause me to vote for politicians with higher ratings."
3. If possible, we find an actual small political race and use the prototype site to perform an experiment.
4. Finally we show the site (and experimental results, if any) to existing, well run organizations that could implement the full idea.

Thus the project has two goals: (1) To convince one or more qualified organizations to implement politician ratings, and (2) To experimentally test the use of politician ratings on a real system.

The Role of Politician Ratings in Society

This document assumes you have read the *Dueling Loops of the Political Powerplace* paper. That paper presents an analysis showing the root cause of change resistance to be a dominant Race to the Bottom Among Politicians. It appears this can be resolved by pushing on the high leverage point of general ability to detect political deception. The paper presents three solution elements for doing this: the Truth Test, Truth Ratings, and Corruption Ratings.

The latter two elements are what we call politician ratings. Corruption Ratings will be needed in the real politician ratings website. However, they would require such a large amount of work and investment that we will omit them for now and do just Truth Ratings.

It's conceivable that in the future politician ratings will cover the positive side of politics, ie how good a job a politician did. But for now politician ratings cover only the negative side of politics. According to the Dueling Loops model this should be enough to make the Race to the Top Among Politicians

go dominant. *Making that loop go dominant and stay dominant is the central role of politician ratings.*

There's another way to look at the role of politician ratings. The average voter does not have the time or skills to evaluate a slate of politicians, or even one politician in depth. The issues have grown far too complex for that. But the larger problem is the rhetoric/falsehoods employed in garnering supporters has grown to be so clever and pervasive that it is impossible for the average voter to rationally figure out who to vote for. The amount of deception has grown so high in most political systems it's comparable to the "fog of war," which has long been known to make sound decision making difficult or close to impossible.

Politician ratings will slice through the "fog of politics" by providing a reliable measure of who is more trustworthy. In most political races this is the single most important factor. "Who can I trust?" is the most important question for the average voter, rather than "Who supports where I stand on issue X?", because the latter questions mostly cancel each other out, there are so many issues. But the "Who can I trust?" question does not cancel itself out, because nearly all voters are asking the same question. If voters knew who they could trust the most, their decisions on Election Day would be ten times as easy to make as they are now, and the results would be ten times as effective.

Society already employs a multitude of ratings. We have product ratings, credit ratings, bond ratings, restaurant ratings, professional competency ratings, hospital ratings, and more. The purpose of every one of them is to allow users of a product or service to make more correct decisions more efficiently. Politician ratings would do the same, in an area that is arguably more important than any other existing rating.

Experiment Overview

The first hypothesis is that in cases where one politician uses more falsehoods than his or her opponent, Truth Ratings will cause voters to more often elect the candidate with the higher Truth Ratings.

The second hypothesis is that in a political system where political falsehoods are common, Truth Ratings will cause such falsehoods to decrease in general.

For Truth Ratings to work they need to be perceived as *unbiased* and *reasonably correct*. They also need to be considered by a sizable fraction of voters, which is estimated to be a minimum of 20%.

This is not a randomized controlled experiment with enough subjects to be statistically significant. Instead, it is a very simple experiment involving one race. The purpose is to quickly and cheaply illustrate how the hypotheses may be true, and to encourage further experimentation and refinement of these ideas.

Here's how a simple experiment is run:

1. First we take past political statements and rate two politicians on them using Truth Ratings. These may be from debates, ads, interviews, speeches, etc. To increase the likelihood of favorable results we should choose a race in which the outcome is expected to be close and one or both politicians are known to be using falsehoods extensively. The ideal race would be one where the politician expected to win is the one employing large amounts of falsehoods to gain an edge.
2. Then we put the Truth Ratings rationale and results on the website and get them published locally, in the paper and such, along with an article(s) about what politician ratings are and why they can lead to better election outcomes. This drives voters and the politicians involved to the site. The articles and site say future statements will also be rated.
3. Then as the months go by we rate further statements, put them on the site, and look for trends. Is less falsehood occurring? If so the Truth Ratings may have worked.
4. The election comes. Did the results favor the candidate with the highest Truth Ratings? If not, was a major issue the reason?

This will not be an easy experiment to run because we will be pioneering a new political process. Ingenuity and quick reactions to unanticipated problems will be required. But if the hypotheses are supported, even weakly, then the experiment can serve to start opening a lot of minds to a new way of thinking. Furthermore, even if the hypotheses are not supported, if the experiment is well run then that will have the same effect.

Avoiding Bias with the Right Process

It is hugely important to minimize bias. It cannot be eliminated, but it can be reduced to the same low amount we see in judicial systems. *If bias is not low the entire project will probably fail.*

A problem is this is a volunteer project. Most of the volunteers will probably be progressives. Right away this biases them toward their favored candidates. Here's the ratings process that may solve the bias problem:

1. We develop a way to *randomly select statements* to rate. They are drawn from the same period of time and of similar importance and type for each candidate. Statements in that pool are selected randomly.
2. The statements are then *identified with a number*. This hides who said it, although some raters will be able to figure that out. The numbers allow "blind" ratings.

3. The statements are then given to the *raters* (at least 3), who rate the statements using *written criteria* they have been trained in. The use of multiple raters is similar to the practice of multiple judges or a jury. The raters *write down their rationale* for arriving at a rating. If necessary the raters can postpone a statement rating until an expert opinion is obtained on related information. It may turn out that we need classes of raters for classes of statements, such as economic, military, general, political, scientific, etc. We will probably need researchers to assist the raters.
4. The rated statements then go to *checkers*, who check them to see that the written criteria for rating has been followed. The checkers look for bias, double check all sources, check the logic in the rationale, etc. The statements are still only identified by a number, so the checkers don't know who rated which statement. If a statement fails the checking step, it is returned to its rater.
5. *The results are examined* for each statement. If the ratings are close, then the average becomes the rating. If not close, then we have a difference of opinion that must be rationally resolved. This is done by the raters conferring, in an identical manner to the way a jury or panel of judges confers when necessary. If the raters cannot agree to within a close range, then that statement is tossed out. It is unratable.
6. The statements for each politician are then *averaged* and that becomes the Truth Rating. An estimated 30 to 100 statements is required. We will later determine the minimum mathematically and for now will assume 30.

The remaining steps are not concerned with bias, but with publishing:

7. The rationales then go through *copy editing*. This reduces the need for the raters to turn out polished prose for the public. Instead, they can focus on quality of ratings.
8. The politician ratings, statements, statement ratings and rationales are then *published* on the website. Simultaneously the ratings are also published in the local paper of record, along with an overview of the results. On the website, users start at politician ratings and can drill down for as much further information as desired.
9. Once a politician has been rated several times, *graphs* showing their ratings over time can be created.

Rating Written Criteria

This has yet to be written. The key strategy is that a statement rating is not the subjective opinion of the rater. *It is an objective conclusion based on evidence.*

The written criteria is expected to be a small booklet, with process steps, checklists, decision making rules, etc. It is used by two types of participants:

raters and checkers. It will be designed so that if followed closely, approximately the same ratings will result, no matter who the raters and checkers are. The variance should be no more than plus or minus 5% for the average of 30 ratings.

The effectiveness of the written criteria can easily be measured by an experiment in itself.

Getting the Ratings Reasonably Correct

This requires considerable expertise. The average person is not qualified to analyze complex political statements and pass judgment on their correctness. This requires high intelligence, training, practice, and vast knowledge of the subject matter. These requirements point to people with a relevant degree and a successful, unbiased, respected career in a field(s) related to the types of statements they are rating. This is much like the requirements for a judge or bond rater.

Finding qualified people could be a problem. We don't yet know if we can enlist such people. At an estimated 60 minutes apiece for the entire process, 30 ratings would take 30 hours. Training in how to do ratings must be added to that, though this only occurs once.

But then again, if we can paint participation as important and as a historic chance to contribute to the future of a better world, maybe we can attract great raters. (Here "raters" includes everyone involved in the process: the statement finders, raters, checkers, researchers, copy editors, etc.)

Who the raters are would be published on the website, with their background and qualifications. This is a high status role. If the average person using the ratings feels the raters are well qualified and unbiased, they will tend to think highly of the ratings.

In the long run the raters would be paid positions. But for the experiment we will need volunteers.

There are similar qualifications for the rest of the participants. These are the people finding the statements, numbering them, giving them to the raters, and then getting the results and calculating the Truth Ratings.

The Visitor Experience on the Website

Ben wrote "I envision the website to be very simple (the average person doesn't spend much time on a website)." This is a good goal. Even though a full politician ratings website will be very large and complex, it must remain conceptually simple for the average visitor. The prototype site has an easier job of this, since it's rating only one political race. But the challenge of perceived simplicity remains.

An established way to design websites is to first create a set of use cases and then design the site to support them. A **use case** is a list of steps a user takes to accomplish a task. Each task is a goal important to the user. First you list the goals/tasks, then you develop the use cases.

These are the key goals the average visitor to the Truth Ratings website is expected to have:

1. Questions, like:
 - What is this website all about?
 - What are Truth Ratings?
 - How are the ratings calculated?
 - How is bias minimized?
 - What is the Truth Test?
 - How can I contact you?
2. What is the rating for a certain politician?
3. How was that rating calculated?
4. What is the rationale behind a rating?
5. I think a rating is wrong. How can I input on that?

The use cases for the above will be developed later. They are simple.

We can get some design ideas from ConsumerReports.org. This is probably the best similar ratings oriented website. Its hallmarks appear to be education in small chunks, great easy to grasp displays of rating data, and easy quick navigation. It projects an air of low bias and trustworthiness.

The ConsumerReports.org site is loaded with images. This adds to its perceived simplicity, readability, friendliness and modern look. We could do the same, and use images associated with candidates and statements. However, there is no strong need to use lots of images, especially if they don't add real value. *It may be best to use few images, because our key content is data and the information behind it.* This would be much like Wikipedia.org.

Here's a starter image policy:

1. No candidate photos. Name only. This avoids the problem of unflattering images, great versus bad photos, and the whole process of getting approved photos. It also contributes to the idea that voters should elect politicians on the basis of something else besides looks.
2. Rating graphs. Once we have multiple ratings for politicians, then we can offer graphs over time of rating trends. These graphs would become the most important images on the site, because they say so much in such a small space. Graphs can be for individual politicians, races, parties, political units like counties, states, countries, and so on. In the experiment we would have just 3 graphs: one for each politician, and one for the race they are in.
3. No images for anything else except to illustrate supporting data for statement rating rationales, such as graphs and photos. We shall see what's best here.

The Participant Experience on the Website

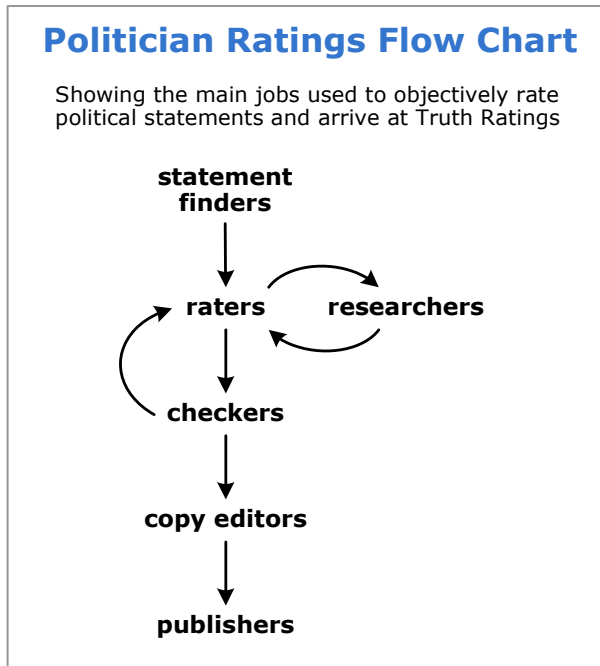
Participants are the raters and those performing the rest of the rating process steps. They have complex use case requirements. How the participants work together may be seen in the diagram. Their use case goals are:

(to be listed later)

This diagram and the rating process described earlier are a first iteration. The total process will evolve and improve over time. Study of how high quality ratings are done in the rest of society should offer tried and true ways to achieve consistent high quality.

Our key challenges are to achieve low bias, high correctness, and high credibility. If we can do that the project will succeed and we won't have to worry much at all about publicity.

Once society discovers it needs politician ratings, it will figure out how to fund them. Something like the above roles will become career jobs.



Website Construction Stages

The site consists of two main parts: pages for *visitors* who use the ratings and supporting material, and pages for *participants* who create the ratings. The use cases will be fairly simple and easy to design for the visitor part.

At a meeting with George and Wilma Turner on June 21, 2008 with the first version of this document, George suggested that we don't need to wait for an actual experiment. We can go ahead and do a rating and learn a lot from it.

For the participant part of the site, we probably need to follow this suggestion. We can manually do an actual mini rating so that we can stabilize the process. Then we can design the participant part. To save time we can do only one politician and just 5 statements. Use case patterns should become clear.

If we do the above then the prototype site would go through these stages:

1. Visitor part only.
2. Visitor and participant parts, based on a small number of statements.
3. Both parts, after we have showed the site to numerous people for feedback.
4. Both parts, based on a real rating of a real race. This would require 60 statements from the experiment. We can do multiple experiments.

To do stage 1 we would need to design the use case for it first. Then we can design the visitor part of the site, and then build it.

To do stage 2 we need to do a small number of ratings, and then use that experience to design the participant use case. Then we can design the participant part of the site, and then build it.

Stage 3 will require serious management of selection of test visitors, getting their feedback, and then using it to incrementally improve the site. Then we do another cycle, etc, until done.

Stage 4 is self-explanatory. By then we will know what we're doing. ☺

Comments

Even though we are building a prototype, for the full effect and to emphasize independence we need to use a new domain rather than make the prototype website part of Thwink.org. The usual approach is a domain name that clearly describes what the site is all about. This points toward something like *PoliticianRatings.org*, which has been reserved in case we need it. Interestingly, *PoliticianRatings.com* is not available. It's a one page website under construction, since at least November 2006. It looks like a dormant project. I've emailed the site.

We need to design in a way to include the behavior of candidates up for reelection in the ratings. For example, if they have made what turned out to be false promises, then that needs to heavily reduce their Truth Rating.

Summary

The purpose of this document is to paint a vision that, if we achieve it, will make a HUGE difference in the course of civilization.

As mentioned earlier our key challenges are to achieve low bias, high correctness, and high credibility. *Thus the emphasis in this website project is not on the user interface of the site, as is the usual case. It's on what's behind it.* That's what will make the difference when we show it to established organizations (like FactCheck.org or Consumers Union) to see if they would be interested in taking the concept further. They can polish the site as they see fit or scrap it altogether. Indeed, we should expect they will do a radical makeover, because they will have different considerations, including scale, budget, a long term plan, and their own unique ideas.

There is, however, no reason not to design the site so well that it is a superlative example of what's possible. My personal dream, and I suspect Ben's and Rob's also, is that the site is so well done and projects its underlying message so well, that voters, journalists, and even a few politicians get excited about this project.

And then the rest, as they say, is history.