The Vote Compass Algorithm

US Presidential Election 2012

1 Introduction

The Vote Compass results section is comprised of multiple elements. The first and most prominent element is the two-dimensional graph, which serves as an abstract representation of the American political landscape. This graph consists of a social and an economic dimension. The second element is a bar graph, which displays a user's level of agreement with each candidate across the 30 key Vote Compass questions. Unlike the two-dimensional graph, the bar graph represents the average distance of a user to each candidate on the issues. These are two different measures of two different concepts.

The consequence of including multiple measures is that there will sometimes be disagreement between the candidate that appears "closest" on the two-dimensional graph and that which appears "closest" on the bar graph. One reason for this is that these graphs are representations of different concepts. It is also because there is no perfect measure of political position, either ideologically or on individual issues. In a public tool of this nature, it is necessary to recognize the trade-off between increased methodological sophistication and the ease with which a method can be understood by the public. The use of multiple measures admits as much. It is an acceptance of the reality that the political world—both among politicians and the public—is complex. It is this complexity that makes politics so lively and contentious, and why successful policies and politics often require great imagination from the public and their political representatives. The purpose of this tool is thus to engage its users to think through this complexity; to encourage those using Vote Compass to learn where candidates stand on the issues and the reasons that they do so; and to raise the level and quality of political information among the public more generally. It is for this reason that we encourage all users to go beyond the results section by clicking on each candidate on the two-dimensional or bar graph to see where the candidates stand on the issues, and to read the documentation that supports the "coding" of these positions.

2 Two-dimensional Graph

The components used to determine a user's position on the two-dimensional graph are the following:

1. A user's responses to the 30 Vote Compass attitudinal and policy-related questions on a scale from 1 to 5.
2. A user's responses concerning the importance of the issues on a scale from 0 to 10.

Each question is defined by whether it is a 'social' or 'economic' question and by the side of the ideological spectrum that responses to the question are on. What each question indicates ideologically is determined theoretically and subsequently checked against data from a poll taken before the launch of the tool. This means that the same response (e.g. 'strongly agree') to one question may differ in ideological direction than the same answer to another question.

User and Candidate Position (Unweighted)

Once a user has answered the 30 Vote Compass questions, his or her position in the political landscape is calculated by averaging the questions x on their given dimension $j \in \{economic, social\}$:

$$position_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}, \tag{1}$$

where n_j is the number of questions for dimension j. The resulting user coordinates are represented by the gray dot labeled 'You' on the two-dimensional graph. Candidate positions are calculated in the same way, given each candidate's answers to the same questions answered by the user. These candidate positions are represented by candidate colors and labels. Clicking on a candidate's logo permits the user to view that candidate's position on each issue and to compare it to his or her answer.

The ellipse surrounding the user's position is a simple measure of reliability. The length of the axis for each dimension j is the standard deviation of the questions x_j:

$$ellipseAxis_j = \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ij} - position_j)^2} \tag{2}$$

User and Candidate Position ("Weighted" by Issue Salience)

Once the user enters the results page, he or she can recalculate the two-dimensional graph based on the issues that are most important to him or her by selecting "Weight your Results." Doing so re-weights the two-dimensional graph by the importance that the user attributed to various issue categories found in the "Importance" section of the questionnaire. Each question is weighted by the priority a user attaches to the category under which a given question is grouped.

Within-dimension Weighting

Within the "Importance" section, the user is invited to evaluate the importance of various categories of issues. Each category is given importance on a scale from 0 to 10. These values define a weight w_{ij} for the question x_{ij} that pertain to each category. When the user chooses to include issue importance on the results page, the graphs are re-weighted both for the user and the candidates.

This weighting is calculated for all questions x with weights w on each dimension j using a weighted average:

$$position_j = \frac{\sum_{i=1}^{n_j} w_{ij} x_{ij}}{\sum_{i=1}^{n_j} w_{ij}} \tag{3}$$

The length of the axes of the ellipse are also re-weighted accordingly:

$$ellipseAxis_j = \sqrt{\frac{\sum_{i=1}^{n_j} w_{ij}(x_{ij} - position_j)^2}{\sum_{i=1}^{n_j} w_{ij}}} \tag{4}$$

Between-dimension Weighting

Before the user chooses to recalculate his or her results by importance, the two dimensions are weighted uniformly. Once values for the importance of each question are included, there will not only be differences in salience within dimensions, but also across them: a user may view social issues as more important than economic ones or vice versa. This is captured in the average level of issue salience that is given by the user for each dimension. The relative importance of a dimension is taken as the average level of salience for one dimension relative to the other. The greater the average salience given for one dimension relative to another, the more the user is assumed to find that dimension salient more generally. The dimensions themselves are re-weighted to account for this difference.

To do this, the less salient dimension is weighted by the ratio of its average salience to the average salience of the more salient dimension. Visually, this has the effect of compressing the dimension that is considered less important to the user, and places the user's coordinate closer to the candidate that is closer to the user on the more salient dimension.

It is important to note that this between-dimension weighting necessarily lowers the maximum value that the candidates' and user's coordinates can take on the re-weighted dimension. The re-weighted graph is therefore a relative measure of ideological placement rather than an absolute one.

Weighting the Candidate Answers

The candidates are weighted, within and between dimensions, by the same salience weights provided by the user. The algorithm places the user closer to the candidate to which the user agrees most on the salient issues and away from those candidates he or she agrees with least. The more salient an issue, the closer the user moves to candidates in agreement with his or her response. If an issue is given a salience value of zero, it is given zero weight, effectively removing it as part of the dimension on which it is situated.

3 Issue Position Bar Graph

Like the two-dimensional graph, the issue-position bar graph has both an unweighted and weighted version. The bar graph measures the absolute distance of the user's issue position to that of each candidate. Once the user indicates the relative importance of the 30 questions, the bar graph is weighted accordingly.

Agreement with Candidate (Unweighted)

To calculate a user's "disagreement" with each candidate, the algorithm takes the sum of the absolute distances of the user's positions x to the positions x_c for those issues of each candidate c:

$$disagreement_c = \sum_{i=1}^{n} \sqrt{(x_i - x_{ic})^2}, \tag{5}$$

where n is the total number of questions. To calculate the relative amount of a user's disagreement with each candidate, the algorithm first determines the maximum

possible distance candidates can be from the user, given the user's answers:

$$maxDisagreement = \sum_{i=1}^{n} \sqrt{(x_i - 3)^2} + 2 \tag{6}$$

This equation centers the scale and takes the absolute value of a user's response to determine its distance from the center. It adds 2, which is the maximum distance a candidate can be from the center. By example, if a user answers '1', a candidate position of '5' is the maximum distance (4) from the user. If a user answers '2', a candidate position of '5' is the maximum distance (3) from the user. If a user answers '3', a candidate position of '1' or '5' is the maximum distance (2) from the user, and so forth. The equation therefore finds the sum of the maximum distance a candidate can be from the user on each of the questions.

The final agreement score with each candidate c is calculated as follows:

$$score_c = \frac{maxDisagreement - disagreement_c}{maxDisagreement} \tag{7}$$

Therefore, if candidate c's positions are perfectly in line with the user for all questions, the user's agreement score with candidate c will be 1. If a candidate's positions are the maximum distance from a user's responses for all questions, the user's agreement score with candidate c will be 0. i.e. For a score of zero, the candidate has issue positions as far as possible from the user, given the user's answers.

Agreement with Candidate (Weighted)

The introduction of weights w for issue salience are included in the calculation as follows:

$$disagreement_c = \sum_{i=1}^{n} w_i \sqrt{(x_i - x_{ic})^2} \tag{8}$$

The calculation for the weighted $maxDisagreement$ follows similarly:

$$maxDisagreement = \sum_{i=1}^{n} w_i (\sqrt{(x_i - 3)^2} + 2) \tag{9}$$

4 Candidate Positions per Question

The elaboration of the Vote Compass questionnaire follows a two-part research process. First, a content analysis is performed on the policy issues that figure most prominently in the platforms and public statements of the main presidential candidates, and in media discourse about national politics. From an initial list of questions, we select those to be included in the final questionnaire based on the questions' ability to differentiate between candidates and amongst voters; their breadth of coverage across multiple policy fields; and their salience in the upcoming election.

Second, candidate positions in the Vote Compass questionnaire are derived from the candidates' publicly-available statements. The Vote Compass research team undertakes a comprehensive review of candidate documents, including election platforms, websites, speeches, press releases, debates, and statements to media, in order to impute an accurate representation of candidates' stances on the policy issues explored in Vote Compass. Preference is accorded to public statements that are recent; come directly from the candidates; and are directly relevant to the policy issue in question. Specifically, public statements are prioritized by date in the following order:

1. Candidates' election platforms
2. Candidates' official policy documents
3. Statements from the candidates (debates, speeches, media); press releases from the candidates
4. Party platforms
5. Other media

Within these guidelines, allowances may be made for statements that most closely represent a candidate's position on the exact phrasing of a particular Vote Compass proposition. This calibration process is followed by a consultation with the candidates themselves. These two steps are described in more detail below.

The Calibration Process

Based on the collected public statements, researchers from the Vote Compass team are assigned to code or calibrate a given candidate's positions on each of the final questions included. To ensure inter-coder reliability, the researchers initially undertake this task separately and subsequently compare results for consistency. As all response categories are presented as Likert-type (or rating) scales, the following guidelines are used in the calibration process:

- **Strongly dis/agree, much less/more, many fewer/more, much harder/easier**
  The candidate clearly emphasizes the issue in question, and does not place any conditions, qualifications, or restrictions on his or her position.
- **Somewhat dis/agree, somewhat less/more, somewhat fewer/more, somewhat harder/easier**
  The candidate does place conditions, qualifications, or restrictions on his or her position; or emphasizes only part of the proposition.
- **Neutral, about the same as now**
  The candidate addresses the issue without consistent argumentation in favor or opposition; defers taking a position; and/or mentions the issue indirectly.

Calibrations on questions pertaining to taxes and spending are based on support for nominal change. In the event that a candidate supports an increase/decrease in taxes or spending that has yet to come into effect, this is still considered support for a nominal change.

To ensure that the results of this process are transparent for users, all candidate positions and supporting public statements (with URLs) are made available in the Vote Compass tool under You vs. Candidate and Candidate vs. Candidate on the results page. This information enables users to compare their own responses to those of the candidates, and to delve more deeply into candidate platforms and public documents.

Consultation with the Parties

Although candidate placements are based primarily on the process explained above, Vote Compass also consults with the candidates themselves as an additional check on our internal research. Candidates are first sent a copy of the Vote Compass questionnaire with accompanying calibrations and supporting documentation, and invited to review their positions on the initial list of questions. Upon receipt of a completed questionnaire from a candidate, Vote Compass then reconciles the candidate's self-placements with the calibrations determined by the research team coders. In the vast majority of cases, the calibrations from the candidate and the Vote Compass research team are in agreement. However, as discrepancies may exist, Vote Compass sends the candidate a reconciliation report outlining the confirmed calibrations and the disputed ones across the final Vote Compass questionnaire. All discrepancies are flagged and justified with the candidate's public statements collected by the research team which support the calibrations proposed by Vote Compass.

The candidate is able to respond to each disputed calibration by clarifying his or her position and providing alternate public statements which support his or her self-

placement on the issue in question. In cases where the candidate provides relevant policy statements which conclusively accord with his or her self-placement, Vote Compass will reposition the candidate on this issue. Where discrepancies are not resolved by this process, the disputed placements are sent for deliberation and final ruling to the Vote Compass Working Group, comprised of foremost scholars of American electoral politics. Candidates are then sent final calibrations for review. They are able to dispute these calibrations and supporting public statements throughout the entire run of Vote Compass. If a candidate's stance on an issue changes or if a candidate wishes to provide additional official documentation not considered during the reconciliation process, we will revisit the appropriate calibration to determine if a change is warranted. Whatever the reason, we encourage candidates to consult with us over the course of the election campaign if necessary. Every effort is made throughout the electoral campaign to ensure the accuracy of candidate calibrations based on their publicly available statements.