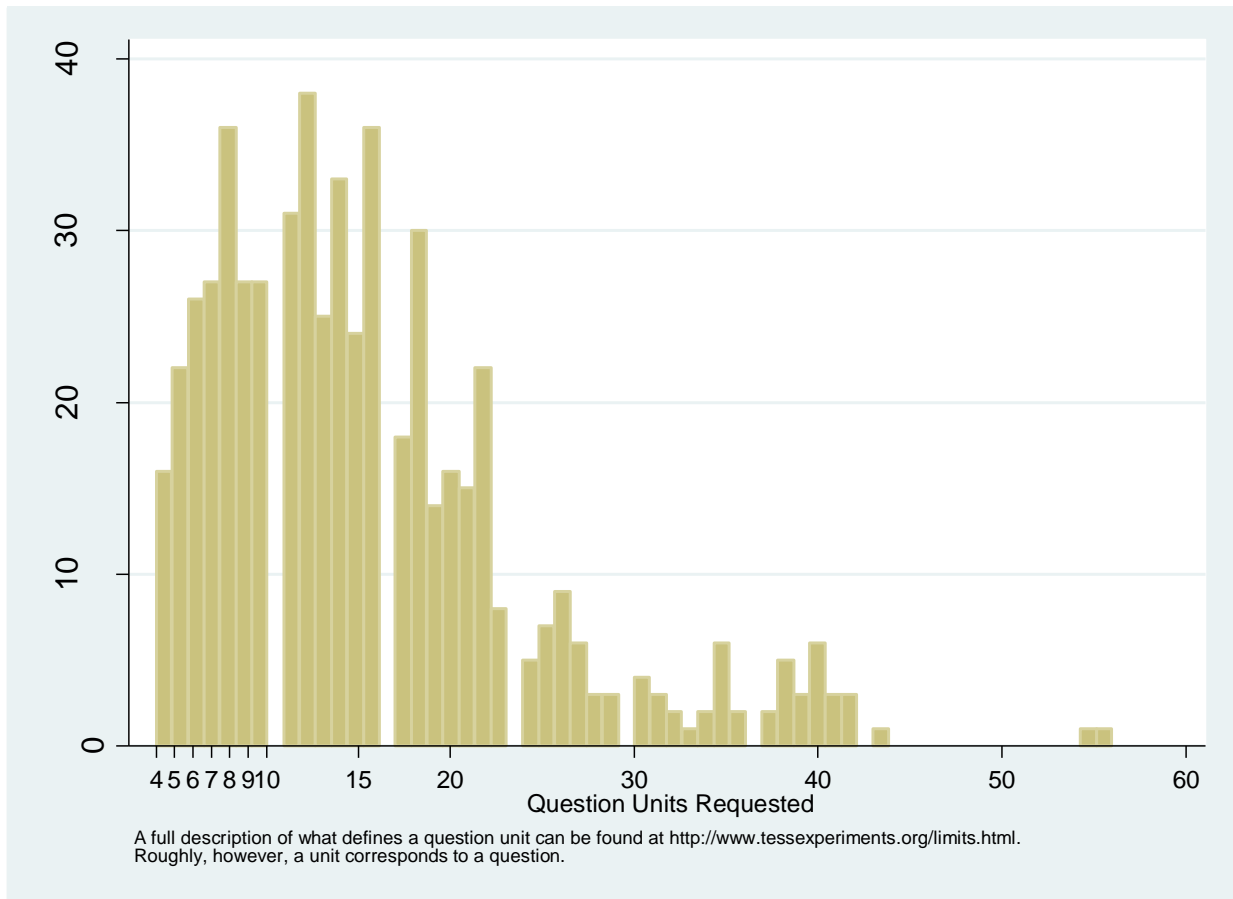


Supplementary Materials

| | |
|---|----|
| Section 1: Number of Question Units Requested for TESS..... | 2 |
| Section 2: Google’s Inferred vs. Reported Demographics..... | 3 |
| Section 3: Description of Canonical Experiments..... | 6 |
| Section 4: Characteristics of Non-Responses across Experimental Groups..... | 8 |
| Section 5: Example of a Survey Wall..... | 9 |
| Section 6: Question by Question Retention Rate (10-Question Survey)..... | 10 |
| Section 7: Rates of Non-Response and “Don’t Know” | 11 |
| Section 8: Limitations in Google Consumer Surveys’ Survey Questions | 12 |

Section 1: Distribution of the Number of Question Units Requested for TESS¹



Section 2: Google's Inferred vs. Reported Demographics

¹ These data is provided by Jamie Druckman, the current PI of the project.

We examine the accuracy of Google’s inferred demographics by replicating the study done by Pew in 2012 on two demographic variables (gender and age) and a third variable income, which was not reported by Pew. Specifically, we asked about 1000 respondents to report their gender, age, and income and then compared these responses to the inferred data provided by GCS. Tables A2.1 to A2.3 below report the comparison between the inferred information and the responses given by the respondents. There are two comparisons on which we focus. First, do reported demographics match inferred ones? Second, since some respondents do not have their demographics inferred by Google, does the distribution of reported demographics differ between those who have inferred demographics and those who do not?

For gender (see Table A2.1), the Pew studies reported that for 75% of respondents inferred gender matched reported gender. However, Pew’s result did not include "I prefer not to answer" as an opt out answer in their response categories -- as generally required by Google when asking respondents sensitive demographic information. When we include this option, however, the percentage of matches decreased to approximately 60%. Nevertheless, when we rescaled our sample by dropping the respondents who chose prefer not to answer option, the accuracy of the inferred gender increased to a comparable level as reported by Pew (around 80%).

Furthermore, we did not find any significant difference in the distribution of reported gender between those respondents whose gender was inferred by Google (column 4) and those who were not (column 5). Even in our original sample, where we include an additional “prefer not to answer” category, a similar trend was also found (although the proportion of those choosing “prefer not to answer” is lower among respondents whose gender was inferred relative to those who were not).

Table A2.1: GCS Inferred vs. Reported Gender

| Gender given in Response | Total Sample (%) | Inferred Gender (%) | | | Unknown (%) |
|-----------------------------------|------------------|---------------------|--------|-------|-------------|
| | | Male | Female | Total | |
| (Pew Sample 2012, N=1056) | | | | | |
| Male | 54 | 79 | 28 | 53 | 58 |
| Female | 46 | 21 | 72 | 47 | 42 |
| (GCS Sample 2013, N=1000) | | | | | |
| Male | 39 | 60 | 16 | 40 | 37 |
| Female | 33 | 13 | 58 | 34 | 26 |
| Prefer not to answer | 28 | 27 | 25 | 26 | 37 |
| (Rescaled GCS Sample 2013, N=718) | | | | | |
| Male | 55 | 82 | 22 | 54 | 59 |
| Female | 45 | 18 | 78 | 46 | 41 |

Figures based on un-weighted data

In terms of age (see Table A2.2), Pew found that from the two samples they surveyed in 2012; there was an average of 44% to 46% of the respondents that reported an age that was in the same category as their inferred age. Again, the opt out choice; “I prefer not to say” was not included in Pew’s response categories. Unlike gender however, the average percentage of the GCS’ respondents choosing a similar age group as their inferred age did not change even when the opt-out option was included in the response categories with an average of 47% choosing a similar age group as their inferred age (thus, we

did not report the rescaled result in this table).² Furthermore, both the Pew studies and ours showed that the largest incongruity between self-reported age and inferred age was observed for middle age respondents (i.e., 35-44) - with only between 23%-43% of these respondents had their age correctly inferred. Finally, we found no significant difference with regards to the distribution of reported age across respondents with and without inferred ages.

Table A2.2: GCS Inferred vs. Reported Age

| Age given in response | Total Sample (%) | Inferred Age (%) | | | | | Total | Unknown (%) |
|---------------------------------------|------------------|------------------|-------|-------|-------|-----|-------|-------------|
| | | 18-24 | 25-34 | 35-44 | 45-64 | 65+ | | |
| (1st Pew Sample 2012, N=1009) | | | | | | | | |
| 18-24 | 20 | 65 | 23 | 12 | 5 | 5 | 18 | 22 |
| 25-34 | 20 | 15 | 39 | 25 | 11 | 5 | 21 | 18 |
| 35-44 | 17 | 10 | 20 | 23 | 16 | 12 | 18 | 15 |
| 45-64 | 31 | 6 | 15 | 34 | 52 | 26 | 30 | 33 |
| 65+ | 13 | 4 | 4 | 7 | 16 | 52 | 13 | 12 |
| (2nd Pew Sample 2012, N=1064) | | | | | | | | |
| 18-24 | 21 | 63 | 21 | 14 | 10 | 11 | 20 | 21 |
| 25-34 | 20 | 11 | 43 | 24 | 10 | 8 | 21 | 17 |
| 35-44 | 17 | 10 | 17 | 30 | 22 | 9 | 17 | 16 |
| 45-54 | 18 | 4 | 7 | 15 | 33 | 18 | 17 | 21 |
| 55+ | 25 | 11 | 12 | 16 | 25 | 54 | 25 | 26 |
| (GCS sample 2013, N=1003) | | | | | | | | |
| 18-24 | 9 | 48 | 17 | 5 | 2 | | 9 | 10 |
| 25-34 | 11 | 11 | 31 | 12 | 3 | | 12 | 11 |
| 35-54 | 32 | 18 | 27 | 54 | 19 | | 33 | 28 |
| 55+ | 29 | 5 | 4 | 16 | 56 | | 29 | 27 |
| Not applicable / Prefer not to answer | 19 | 18 | 21 | 14 | 19 | | 18 | 25 |

Figures based on un-weighted data

Besides replicating the two inferred demographics that were also studied by Pew, we also assessed the accuracy of GCS' inferred income variable - by asking respondents to report their income category (see Table A2.3).³ The results showed that GCS' inferred income is far less consistent with self-reported income than was the case for gender and age. Specifically, we found that only 23.5% of respondents chose an income category that was similar to the one that GCS inferred. Further, contradictions between what respondents report and what GCS inferred were prevalent across all income categories. Finally, unlike the other two demographic characteristics, there was a somewhat different distribution of self-reported income between respondents for which Google was able to infer their income and those for which they could not.

Table A2.3: GCS Inferred vs. Reported Income

² GCS limits the number of response categories to five but infer age in six categories. Pew built two separate samples with two different age categories and collapsed different age categories for each. We only provided four age categories with "Not Applicable/Prefer not to say" being the fifth option. We collapsed the inferred age according to the age categories in our response option.

³ Similar to our previous assessment of age, we collapsed the inferred income into five categories in order to accommodate the number of response categories that GCS allowed. We collapse income group of \$100,000-\$149,999 and \$150,000+ into one category.

| Income given in response | Total Sample (%) | Inferred Income (%) | | | | | Total | Unknown (%) |
|--------------------------|------------------|---------------------|-------------------|-------------------|-------------------|------------|-------|-------------|
| | | \$0-\$24,999 | \$25,000-\$49,999 | \$50,000-\$74,999 | \$75,000-\$99,999 | \$100,000+ | | |
| (GCS Study 2013, N=1006) | | | | | | | | |
| \$0-\$24,999 | 34 | 34 | 35 | 32 | 31 | 50 | 34 | 33 |
| \$25,000-\$49,999 | 18 | 24 | 21 | 13 | 11 | 8 | 18 | 22 |
| \$50,000-\$74,999 | 18 | 17 | 19 | 17 | 13 | 0 | 18 | 22 |
| \$75,000-\$99,999 | 10 | 13 | 8 | 15 | 13 | 17 | 11 | 0 |
| \$100,000+ | 19 | 11 | 17 | 24 | 31 | 25 | 19 | 22 |

Figures based on un-weighted data

Overall, GCS' inferred demographics matched respondents' self-reports in about 74% of cases for gender, 40% for age, and 20% for income.

Section 3: Description of Canonical Experiments

3.1 Welfare Experiment

For the 1-question survey, we simply ask respondents whether too much, too little, or about the right amount were being spent on either “welfare” or “assistance to the poor.”

For the 10-question survey, we ask the following questions:

1. What is your age?
2. What is your gender?
3. Do you think that the government is spending too much, too little, or about the right amount of money in [assisting the poor/welfare]?
4. Which of the following policy areas did we ask about in the previous question?
5. Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?
6. When it comes to politics, would you describe yourself as liberal, conservative, or moderate?
7. What is your highest educational attainment?
8. What is your annual income?
9. Please select 'red' and 'green' among the alternatives below, no matter what your favorite color is.
10. Statistics show that African Americans have, on average, lower incomes than other groups. Which of the following are important causes of this difference?

3.2 Asian Disease Experiment

In the “Asian Disease” experiment, all respondents were given the following scenario:

Imagine that your country is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

They were then given either one of the two following conditions:

Condition 1, Lives Saved: If Program A is adopted, 200 people will be saved. If Program B is adopted, there is one-third probability that 600 people will be saved, and two-third probability that no people will be saved.

Condition 2, Lives Lost: If Program A is adopted, 400 people will die. If Program B is adopted there is one-third probability that nobody will die, and two-third probability that 600 people will die.

In each of these two conditions, respondents were asked to choose one of the two programs.

3.3 List Experiments

We replicated two different “list experiments” (often called the “Item Count Technique”) that are designed to allow researchers to get more accurate responses to questions in which the respondent has an incentive to over or underreport (or otherwise shade the truth).

The first replication is the list experiment conducted by Janus (2010). Both the treatment and control questions offer the same four non-sensitive items to choose from:⁴

1. *The way gasoline prices keep going up.*
2. *Professional athletes getting million dollar-plus salaries.*
3. *Requiring seat belts to be used when driving.*
4. *Large corporations polluting the environment.*

In the treatment condition however, respondents were shown an additional (sensitive) item:

5. *Cutting off immigration to the United States.*

Respondents were only asked to say how many statements that they oppose, not which ones. The hope is that this eliminates the respondent’s concern about giving the socially desirable answer and allows the person to reply more honestly. Because the two questions are randomly answered, the mean number of “oppose” responses to the first four statements should be the same for both the baseline and test groups; therefore, any increase in the mean number of opposed items in the treatment group must be attributed to the “cutting off immigration to the United States” statement. We also replicate a similar list experiment in GCS by Streb et.al (2008) that examines attitudes about a female president. In this experiment, both groups received similar non-sensitive items as the ones given in the immigration experiment while the treatment group receives the sensitive item: "A woman serving as president".

⁴ The order of the statements was randomized.

Section 4: Characteristics of Non-Responses across Experimental Groups

Table A4.1 shows that balance on inferred demographics is achieved between the control and treatment groups within the samples that chose to opt out. Note that the overall non-response rates for treatment and control groups are given in the last column.

| Experiment | Gender (%) | | Age (%) | | | | | | Region (%) | | | | Total (%) |
|--------------------------------|------------|------|---------|-------|-------|-------|-------|-----|------------|------------|-------|------|-----------|
| | Female | Male | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65+ | Mid-west | North-east | South | West | |
| Question Wording | | | | | | | | | | | | | |
| Control | 43 | 57 | 18 | 17 | 15 | 18 | 17 | 15 | 22 | 21 | 34 | 24 | 70.7 |
| Treatment | 42 | 58 | 18 | 17 | 15 | 18 | 17 | 15 | 22 | 20 | 34 | 24 | 70.4 |
| Asian Disease | | | | | | | | | | | | | |
| Control | 43 | 57 | 17 | 17 | 16 | 17 | 19 | 14 | 21 | 19 | 36 | 25 | 74.4 |
| Treatment | 42 | 58 | 17 | 18 | 16 | 16 | 19 | 15 | 20 | 20 | 35 | 25 | 74.7 |
| List (Female President) | | | | | | | | | | | | | |
| Control | 43 | 57 | 17 | 17 | 16 | 15 | 21 | 15 | 21 | 21 | 33 | 27 | 74.4 |
| Treatment | 45 | 55 | 16 | 18 | 15 | 17 | 19 | 15 | 21 | 21 | 32 | 27 | 74.1 |
| List (Immigration) | | | | | | | | | | | | | |
| Control | 45 | 55 | 16 | 18 | 16 | 16 | 20 | 15 | 20 | 21 | 34 | 26 | 76.2 |
| Treatment | 47 | 53 | 15 | 18 | 15 | 19 | 18 | 17 | 21 | 21 | 33 | 27 | 74.9 |

Figures may not add to 100% because of rounding

Section 5: Example of a Survey Wall

Fair Use Digital Circulation Strategy Information Overload

The Work of Art in the Age of Mechanical Reproduction

Matthew Dodd from the January 16, 2015 issue

Jurgen Habermas R&D Android cops beat The Weekender mathewi Tim Carmody attracting young readers tweets, collaboration tags the medium is the message blog plagiarism horse-race coverage advertising the other longer Book Review. ...



Privacy put the paper to bed Fuego news.me photo source: proimos flickr

Please complete the following survey to access this premium content.

Do you think that we are spending too much, too little, or about the right amount of money on welfare?

- Too Little
- Too Much
- About Right
- Don't Know

OR

[Show me a different question](#)

[Skip survey](#)

Google

[INFO](#) [PRIVACY](#)

Popular on The Daily Globe

MOST EMAILED

MOST VIEWED

OPINION

[A Line in the Sand Against Rate Hikes](#)

ENTERTAINMENT

[A-List Guide to Oscar Parties](#)

THE BUZZ

[Memorable Quotes from 2012](#)

SPORTS VIDEO

[LeBron James Youngest to Score 20000](#)

HOME & GARDEN

[How to Plant a Winter Garden](#)

Section 6: Question by Question Retention Rate (10-Question Survey)

Table A6.1 shows that the initial response rate is already very low (i.e. 12%) for both groups. In the end, only 63% of the respondents who answer the first question completed the entire survey.

Table A6.1: Retention Rate by Question (10-Question Welfare Experiment)

| Order | Question | Control | | Treatment | |
|------------------------------|-------------------------|-------------|--------------|-------------|--------------|
| | | Impressions | % Continuing | Impressions | % Continuing |
| 1 | Age | 7584 | 12% | 8440 | 12% |
| 2 | Gender | 931 | 92% | 1044 | 92% |
| 3 | Treatment | 853 | 93% | 962 | 94% |
| 4 | Manipulation Check | 794 | 90% | 900 | 86% |
| 5 | Party ID | 713 | 94% | 778 | 96% |
| 6 | Ideology | 673 | 97% | 750 | 97% |
| 7 | Education | 655 | 98% | 731 | 99% |
| 8 | Income | 640 | 96% | 721 | 94% |
| 9 | Trap Question | 615 | 98% | 677 | 97% |
| 10 | Attitudes toward Blacks | 602 | 83% | 654 | 80% |
| Completed | | 502 | | 524 | |
| Overall Response rate | | 7% | | 6% | |

Section 7: Rates of Non-Response and “Don’t Know”

Table A7.1 shows that there are similar proportions of respondents who choose to opt-out, engage the question (by giving substantive answer), and select “don’t know” across control and treatment groups for the five experiments we conducted.

| | Framing (1-qn) | | Framing (10-qn) | | Asian Disease | | List (Female President) | | List (Immigration) | |
|-------------------------|----------------|-----------|-----------------|-----------|---------------|-----------|-------------------------|-----------|--------------------|-----------|
| | Control | Treatment | Control | Treatment | Control | Treatment | Control | Treatment | Control | Treatment |
| Answer (%)* | 64 | 64 | 75 | 75 | 59 | 58 | DK not provided | | DK not provided | |
| DK (%)* | 36 | 36 | 25 | 25 | 41 | 42 | DK not provided | | DK not provided | |
| Total Response Rate (%) | 29 | 30 | 10** | 11** | 26 | 25 | 26 | 26 | 24 | 25 |
| Non-Response (%) | 71 | 70 | 90 | 89 | 74 | 75 | 74 | 74 | 76 | 75 |

* The values here refer to the percentage of respondents who do not opt-out of the survey

**As the main treatment question appears as the third out of ten questions, this number is the proportion of respondents who have answered the first two c

Section 8: Limitations in Google Consumer Surveys' Survey Questions

Below are some restrictions that GCS imposes in constructing survey questions (as of April 2016).

- a) *Number of response categories*: GCS tries to ensure the survey is simple by restricting response categories to a maximum of seven and an option of adding one open-ended category - i.e., "Other (please specify)".
- b) *Number of characters/words*: To keep the question short, GCS recommends the question length to be 125 characters and sets the maximum limit to 175 characters.
- c) *Censored questions and populations*: GCS places restrictions on sensitive demographic information by prohibiting researchers from asking respondents for their age, gender, ethnicity, religion, and immigration status. Researchers can only ask these questions if the response choices include "I prefer not to say" as the opt-out answer. Also, questions may not directly or indirectly target respondents under 18 years old or include any age category under 18 in the answer choices.
- d) *Don't know and "opt-out" option*: As we discussed previously, besides regulating the content and the target respondents of the question, GCS also requires an option for the respondents to not answer the question by either clicking an "I don't know, show me another question" link, or by sharing the premium content that they intend to read in their social media pages (see Figure 3 above). Essentially, this just ensures that all respondents have some path that allows them to skip a given question yet still (ultimately) access their desired content.

Working within the Word Constraint

One of the constraints of using GCS for survey experiments is the word limit for questions and response sets. For many kinds of survey experiments of interest to social scientists, these word limits will be too constraining. Indeed, in many of the canonical experiments we conducted, we could not faithfully implement the experiments under these word limits (even after a great deal of work streamlining wording).

Thus, we devised one way to work around this limitation by using the "picture technique." Specifically, we captured the necessary text in a picture format that was then inserted into the question as a screen image. Below is an example of doing this for the response set in a list experiment. Similar pictures can also be utilized for longer questions or textual vignettes, as is necessary in many survey experiments.

Below are the five things that sometimes people oppose or are against. HOW MANY of them do you oppose?

- Requiring seat belts to be used when driving
- Cutting off immigration to the United States.
- The way gasoline prices keep going up.
- Professional athletes getting million dollar-plus salaries.
- Large corporations polluting the environment

The preamble and question fit within the word limit, but the listed items do not and so we insert them as a graphic in the question itself.

Of course, while one can use this technique to circumvent the word limit, it is imposed because of real concerns that respondents in a “survey wall” environment are less motivated than those in other survey settings. Thus, researchers should carefully consider the difficulty of the task they are asking the respondents to do.