EoA study Pre-Analysis plan

February 5, 2016

Design

500 mTurk workers will be asked to estimate the probability of two events - the outcome of the 2016 Super Bowl, and the outcome of the 2016 Republican New Hampshire Primary. They will then write an argument for one of the sides, and re-estimate the same probabilities. In all treatments participants will give probability estimates for both events, before and after the treatments.

The treatments will be randomized along two factors, in a 2x2 between-subjects design (plus a fifth control condition). One factor will vary the target event - half the subjects will argue a side for the Super Bowl, and the other half will argue a side for the GOP primary. The other randomized factor is the stance of the argument they are asked to write.

Some will be assigned to an *advocate* stance, in which they will put forward their most persuasive argument **in favor** of the outcome that they think is most likely. Others will be assigned to an *opposite* stance, in which they will put forward their most persuasive argument **against** the outcome that they think is most likely. A control condition will have participants write a description of a shopping trip, without any argument stance at all.

Analysis Plan

Exclusion Criteria

We plan to exclude participants from our analyses based on the following criteria:

- did not complete the full survey
- failed the initial attention check (however, the survey is designed to reject those who fail the attention check before they encounter any of the key dependent measures, and they will not be included in subject counts)
- unwilling to write more than 25 words for the initial writing assignment
- repeat participation from the same IP address

Outcomes

Our primary dependent variable is the change in probability estimates within the study (i.e. the difference in estimates before and after arguing). The accuracy of these estimates will be determined using an objective measure of the likelihood of the different outcomes in each event. This measure will be decided by using the official forecasts on fivethirtyeight.com, timestamped at 12pm EST on Friday, February 5, 2016. Here are the links to those pages:

http://projects.fivethirtyeight.com/election-2016/primary-forecast/new-hampshire-republican/
http://projects.fivethirtyeight.com/2015-nfl-predictions/

Note that the primary forecast page includes two forecasting models - we will be taking the “polls-plus” model as our objective standard, since it is more comprehensive.

Treatment Effects

We do not make any predictions about the main effect of event, so we will collapse across this factor (after controlling for baseline levels of confidence for each event). The remaining predictions all concern the change in estimates based on condition, separately for both the target event (i.e. the one they were assigned to argue) and the non-target event (i.e. the one they were not assigned to argue).

The first three hypotheses describe the direct effect of arguing, on the target event.

H1: the *advocate* condition will increase the participants' estimated likelihood that the outcome of the target event they believe is most probable will occur, compared to control
H2: the *opposite* condition will decrease the participants' estimated likelihood that the outcome of the target event they believe is most probable will occur, compared to control
H3: the *advocate* condition increase the participants' estimated likelihood that the outcome of the target event they believe is most probable will occur, compared to the *opposite* condition.

The next three hypotheses describe the indirect effect of arguing, on the non-target event.

H4: the *advocate* condition will increase the participants' estimated likelihood that the outcome of the non-target event they believe is most probable will occur, compared to control
H5: the *opposite* condition will decrease the participants' estimated likelihood that the outcome of the non-target event they believe is most probable will occur, compared to control
H6: the *advocate* condition will increase the participants' estimated likelihood that the outcome of the non-target event they believe is most probable will occur, compared to the *opposite* condition.

Moderators

There are three moderators of the main effect that we wish to investigate.

First, we predict that the effects of argument stance will be strongest among people who have a personal preference for the outcome which they believed was most likely. This will be assessed in a separate question in the survey, to determine which outcome would be most desirable in each event.

Second, we predict that the effects of argument stance will be strongest among people who place more importance on the outcome of each event. Importance will be assessed in separate Likert-scale questions in the survey, to determine how invested each participant is in the outcome of each event.

Third, we predict that the effects of argument stance will be strongest among people who have the most domain knowledge for the event. Domain knowledge will be assessed objectively, using a short four-item factual quiz for each event.

Full List of Measures

Probability Estimates

Both events have two complementary and mutually exclusive outcomes. The focal outcome of each estimate will be determined based on which outcome the participant thinks is most likely, determined by an initial question.

On Sunday, the Super Bowl will be played between two teams (Denver Broncos and Carolina Panthers). Which of these two outcomes do you think is the most likely?
[forced choice: "Denver Broncos will win the game" vs. "Carolina Panthers will win the game"]

On Tuesday, a Republican Primary Election will be held in New Hampshire, and Donald Trump is one of the candidates. Which of these two outcomes do you think is the most likely?
[forced choice: "Donald Trump will win the NH Primary" vs. "Donald Trump will lose the NH Primary"]

The probability questions will be estimated twice - at the beginning and end of the survey. Participants will give their responses using a slider, that ranges from 0% to 100% probability. The wording will reflect the choice of focal outcome from the questions above, and will be consistent throughout the study.

What is your best estimate of the percent probability that the [Carolina Panthers/Denver Broncos] will win the Super Bowl next week?

What is your best estimate of the percent probability that Donald Trump will [win/lose] the New Hampshire Primary next week?

Outcome Preferences
The Denver Broncos and the Carolina Panthers will play in the Super Bowl next week. Do you have a personal preference for the outcome of the game?

[strong DEN preference; moderate DEN preference; weak DEN preference; no preference; weak CAR preference; moderate CAR preference; strong CAR preference]

The Republican Primary in New Hampshire will happen next week. Do you have a personal preference for the outcome of the primary?

[strong preference for Trump win; moderate preference for Trump win; weak preference for Trump win; no preference; weak preference for Trump loss; moderate preference for Trump loss; strong preference for Trump loss]

Weight guessing

Participants will be shown a picture of a young man and asked to guess his weight. Then they will be asked, "How confident are you that your guess is within 5 pounds of his actual weight?" (with a slider going from 0 to 100%). This question is included to test how far, exactly, the effect of advocacy extends.
(He actually weighs 190.6 pounds.)

Factual Quiz

Each quiz will be scored out of 4, to give an independent measure of domain knowledge for each event. Answers will be given as open-ended text, and judged by a research assistant blind to condition (minor spelling mistakes will be tolerated).

- *which team did the Denver Broncos beat in the AFC championship game to make it to the super bowl?*
- *which team did the Carolina Panthers beat in the NFC championship game to make it to the super bowl?*
- *who is the starting quarterback for the Denver Broncos?*
- *who is the starting quarterback for the Carolina Panthers?*

- *who is the current vice-president?*
- *who is the current speaker of the house of representatives?*
- *who won the New Hampshire Primary Election for the GOP in 2008?*
- *who won the New Hampshire Primary Election for the GOP in 2012?*

Event Importance
Participants will respond to these questions using a 1-7 likert scale, anchored using the stems "not at all important" and "extremely important"

- how important is the outcome of the Republican primary for you, personally?
- how important is the outcome of the Super Bowl for you, personally?

Motivation for Research Questions

This research is designed to understand the role of argumentation in overconfidence. That is, previous work has shown that certain kinds of position-taking can either increase or decrease confidence during the course of a single study, even when no new information about the position is given to the participant. Furthermore, some recent work suggests that these effects can generalize outside of the domain in which the argument is made. Our work is designed to replicate those basic findings (i.e. within-domain increases/decreases in confidence from advocate/opposite argumentation) and also see whether the effects generalize to a different domain.

This research tests the basic determinants of overconfidence - in particular, whether over-precision is increased by argumentation in favor of the focal hypothesis, and/or decreased by argumentation against the focal hypothesis. There is also an important application of these results, especially if the *opposite* condition decreases overconfidence in the non-target event. While many interventions have been designed to better-calibrate specific forecasts, these interventions typically do not have any effect on subsequent judgments, and forecasters are no better . This research would test whether forecasters who are given a debiasing tool (i.e. the *opposite* condition) will spontaneously use that tool in a subsequent judgment, even when that

tool is not prompted. Such a result would provide new hope for more scalable, long-lasting interventions to reduce overconfidence.

References

Cheatham, L., & Tormala, Z. L. (2015). Attitude Certainty and Attitudinal Advocacy: The Unique Roles of Clarity and Correctness. *Personality and Social Psychology Bulletin, 41(11),* 1537-1550.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind improving individual judgments with dialectical bootstrapping. *Psychological Science, 20(2),* 231-237.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: a corrective strategy for social judgment. *Journal of personality and social psychology, 47(6),* 1231.

Minson, J. Skowronek, S. & Chen, F. (2015). The Advocacy Mindset: Arguing boosts confidence in…. everything. *Presentation at the Annual Meeting of the Society for Judgment and Decision Making, Chicago, IL.*

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review, 115(2),* 502.