Preregistration

How to sway voters? Part 6

Ralf H. J. M. Kurvers[1], Uri Hertz[2], Marta P Balode[1], Bertrand Jayles[1], Bahador Bahrami[1,3]

[1] Max Planck Institute for Human Development (Berlin), Center for Adaptive Rationality

[2] University of Haifa, Department of Cognitive Sciences

[3] Ludwig-Maximilians Universität München, Faculty of Psychology and Educational Siences, General and Experimental Psychology

10. February 2020

Study Information

Research questions

RQ1: Can a dishonest advisor successfully draw the attention of a client with incentives?

RQ2: How is the ability of a dishonest advisor to draw the attention of a client affected by the level of uncertainty in the environment with incentives?

Hypotheses

H1 (RQ1): We expect that also with incentives a dishonest advisor is better able to draw the attention of a client than an honest advisor.

H2 (RQ2): We expect that also with incentives a dishonest advisor is better able to draw the attention of a client when there is high uncertainty (i.e., when the outcome of a trial is difficult to predict).

Design Plan

Existing data

Registration prior to creation of data. As of the date of submission of this research plan for preregistration, the data have not yet been collected, created or realized.

Study design

This preregistration is part 6 of the project "How to sway voters". Part 1 of this project can be found here and for full experimental details we will refer to this document. Data collection of part 1 has been finished, and in this part 6 we plan to investigate whether the two main results of part 1 ((i) the dishonest advisor can gain more influence than the honest advisor, and (ii) the dishonest advisor is better at gaining influence under high uncertainty) can be replicated when we incentivize participants for correct outcomes.

This study will use an experimental paradigm similar to the one in Hertz et al. (2017). In brief, there will be three agents in the experiment: two advisors and one client. On each trial, the client (human) has to make a decision which of the two advisors (algorithms) to follow. The advisors observe part of the available evidence, and then decide how to communicate this evidence to the client. The client does not observe any evidence and thus has to rely on the advisors.

Each trial consists of the following steps:

- A rack of 100 balls is filled with a mix of white and black balls.
- Both advisors observe the same (randomly sampled) 75 balls from the rack. (The value of 75 adds a small amount of noise to the signal; that is, not all evidence is available).
- The client selects one of the two advisors to follow.
- Both advisors communicate their advice to the client. The advice consists of (i) colour (black or white), and (ii) level of confidence (1-5 scale).
- One ball from the rack is randomly drawn. If the colour advice of the selected adviser matches (/does not match) the colour of the drawn ball, the client wins (/loses).

In this version of the experiment, the two advisors will be algorithms and play two predefined strategies:

- Honest Advisor (HA) The HA reports the colour and confidence honestly: if the majority of observed balls are white (/black), the HA reports white (/black) to the client. The HA also reports confidence honestly, using the following linear mapping: majority of one colour: 50-60%: Confidence (CF) = 1. 60-70%: CF = 2. 70-80%: CF = 3. 80-90%: CF = 4. 90-100%: CF = 5. Additionally, we add a small amount of noise: in 20% of cases the CF level was increased with one unit; and in 20% of cases lowered with one unit (provided this was possible).

- Dishonest Advisor (DA) When selected, the DA performs the same strategy as the HA, thus reporting the evidence honestly. When not selected DA's strategy depends on the strength of the evidence:

- if the evidence is strong (> 75% of observed balls are of the same colour), DA performs the same strategy as the HA.
- if the evidence is weak (<= 75% of observed balls are of the same colour), DA reports the colour of the **minority** of the balls to the client with a confidence level 2, 3 or 4 (all values equally likely).

The **client** is a human player and will at the start of each trial select one of the two advisors to follow. The experiment runs for 20 trials.

We plan four treatments, varying the uncertainty in the environment. Specifically, we manipulate the ratio of black vs. white balls between treatments:

1. 25% of trials: 90 balls of one colour. 75% of trials: 50 balls of each colour.
2. 25% of trials: 90 balls of one colour. 75% of trials: 60 balls of one colour.
3. 25% of trials: 90 balls of one colour. 75% of trials: 70 balls of one colour.
4. 25% of trials: 90 balls of one colour. 75% of trials: 80 balls of one colour.

From treatment 1 to 4 the level of uncertainty decreases. That is, the trials become increasingly easier. We have specific predictions based on the results of part 1.

Results part 1 In part 1 of this series of studies, we already performed the same experiment, but not incentivizing participants for correct choices. Fig. 1 shows the results of part 1. As can be observed, we find that, as predicted, the **DA** is able to gain a higher influence than the **HA**, but only under conditions of high uncertainty. In this part 6,

we conduct a new study to see if we can replicate this key result when incentivizing participants for correct choices.

Randomization

In our planned study, the **HA** and **DA** will appear either on the left or right side of the screen and this will be counterbalanced between participants. That is, approximately half of the participants will experience the HA on the left side, and the other half will experience the HA on the right side of the screen.

Blinding

In our planned study, participants in all four treatments will receive the same instructions. Participants are unaware of the exact treatment they are in, and they are unaware that there are different treatments.

Data collection procedures

The data will be collected using an online study implemented in the Lioness platform. Participants will be recruited over *Amazon Mechinal Turk* (https://www.mturk.com/). The study will take approximately 15 minutes. Participants who complete the study receive 3 dollar compensation plus a bonus payment. The bonus payment depends on the number of correct lottery outcomes. For each correct outcome a participant receives 10 dollar cent. Bonus payment thus ranges from 0 (no correct lottery outcomes) to 2 dollar (all lottery outcomes correct). Note that this incentivization is the crucial difference with part 1 of this study.

To ensure that participants have understood the instructions before starting the experiment, a series of comprehension questions is asked after the instructions. Only participants who correctly answer all comprehension questions can start the experiment. Participants can go back to read the instructions during the comprehension questions, but can submit their answers maximally five times (to avoid participants who try all possible combinations of answers).

Sample size and stopping rule

In each of the four treatments we plan to collect data for 40 participants. A successful completion means that a participant started and finished the experiment. Participants who quit the experiment prior to completion will be excluded from the analysis. The final sample size might be slightly higher than 40 in a treatment because we collect data on MTurk in badges, and once participants started (but the maximum sample size is already obtained) they can complete the experiment. We

will include these additional data points as well to maximize statistical test power.

Measured variables

The key variable of interest is the *choice* of the clients. Additionally, the following demographic variables will be elicited at the beginning of the study:

1. Age.
2. Gender: female, male, other, do not want to report.
3. Education: basic, high school, college, postgraduate.

Data exclusion

Participants who did not complete the entire experiment will be excluded from the analysis.

References

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02314-5

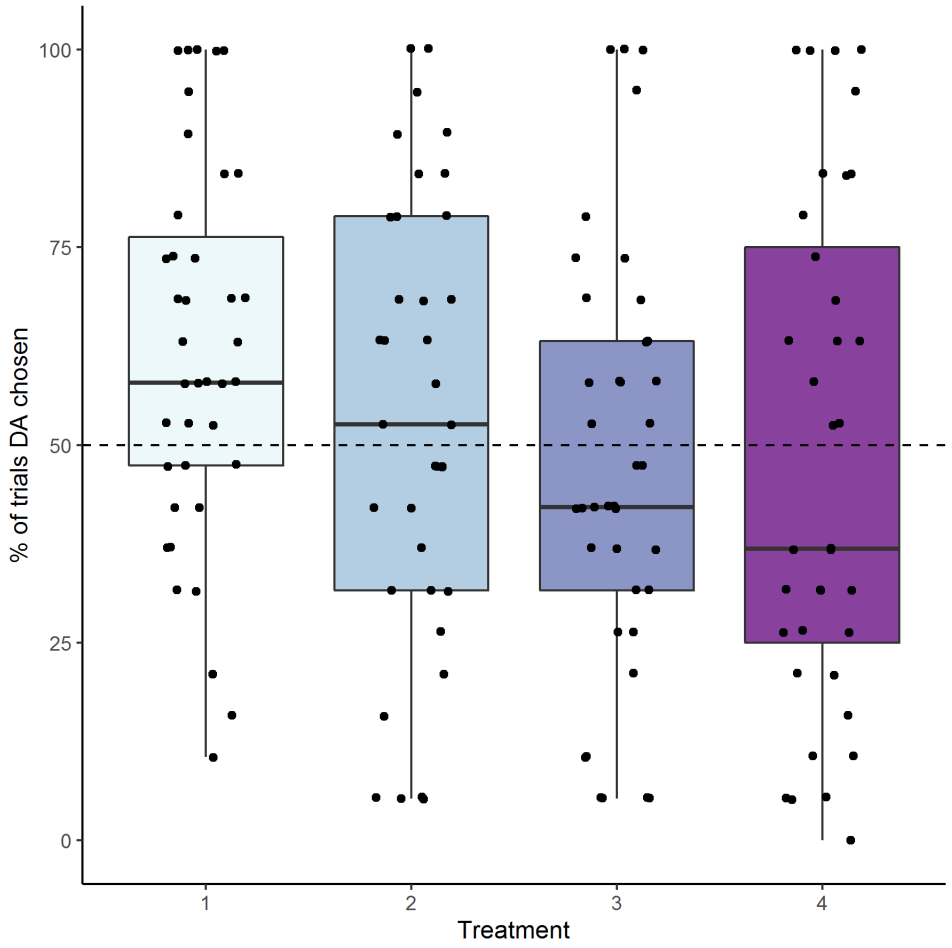

Figure 1: Results of part 1. At the highest level of uncertainty (i.e., treatment = 1), the DA was more successful in being selected than the HA. The lower the uncertainty, the less successful the DA became. Each dot represents the mean choice of one participant. Box plots show the median plus interquartile range of each treatment. Horizontal dashed line indicates chance level.