**Crowdsourcing Data Analysis 2, Phase 2:**
**Explaining Variability in Analyses and Results**

Interested in how analytic choices affect research results? Interested in the role of gender in scientific debates? Do you know how to analyse data using R? Join us as an analyst and co-author for the second phase of our project crowdsourcing the analysis of a dataset on gender, status, and science.

We are employing the new approach of crowdsourcing data analysis, in which many independent analysts are recruited to test the same hypotheses on the same data set. Our first crowdsourcing data analysis initiative examined whether soccer referees give more red cards to dark skin toned than light-skin toned players (Silberzahn et al., under review; see project page on the Open Science Framework at https://osf.io/gvm2z/). The outcome was striking: although approximately two-thirds of teams obtained a significant effect in the expected direction, estimated effect sizes ranged from moderately large to practically nil.

In this second project "Crowdsourcing Data Analysis 2: Science, Gender, and Status" we are in the process of crowdsourcing the analysis of a dataset on intellectual conversations. You can become a co-author on the project by conducting analyses to test two hypotheses regarding the roles of the speaker's status and gender in debates between scientists.

Critically, in this new phase of Crowdsourcing Data Analysis 2 we are trying to pinpoint exactly WHY analytic choices have such a profound effect on research results. We are recruiting scientists who analyse their data using R and are willing to use our new "Data Explained" platform to carefully track their analytic decisions in real time. Using Data Explained, we hope to identify the factors that play a role in data analysis variability. You can see a video tutorial for the platform here: **https://www.youtube.com/watch?v=UVNIFJaeNwI**

**What is the Dataset?**

- The data comes from Edge.org, a platform for intellectual discussion and debate
- The more than 700 contributors are chosen by Edge based on their creative work and include Daniel Kahneman, Marissa Meyer, Craig Venter, and many other academics as well as writers, entrepreneurs, business leaders, and more
- Each row is one comment in a conversation. There are approximately 7,600 rows and 3.8 million words. We also have information about the contributors.

**What are the two hypotheses?**

- Hypothesis 1: "A woman's tendency to participate actively in the conversation correlates positively with the number of females in the discussion."
- Hypothesis 2: "Higher status participants are more verbose than are lower status participants."
- For those less familiar with text analysis, the following resources could be helpful: https://discovertext.com/, http://www.liwc.net/, http://www.uclassify.com/

**Information for Collaborators**

- Project participants must work alone for the analysis phase of the project so that we can track your decisions using the Data Explained platform. *During the analysis phase of the project, you should conduct your analysis independently, without collaborating or corresponding regarding your analysis with other project participants.* At a later stage in the project (see timeline below) you will be able to discuss your analyses with each other and provide each other feedback.
- Every person who completes and submits his or her analyses using the Data Explained platform, and submits his or her final analysis report within the stated timeframe will be an author on the final paper (listed in alphabetical order after the coordination team and before the senior and last author).
- Each project contribution must include: (1) the code for the analysis and specification of analysis

package required to execute the analysis, (2) a description of the rationale for the analysis strategy via Data Explained, (3) a complete written summary of the analysis strategy, and (4) a description of the result including specification of the effect estimate in effect size units (d, r, $R^2$ or odds ratio) and confidence interval.

- We are looking for colleagues with a wide range of expertise to participate in this crowdsourcing project, including researchers interested in text analysis, time, gender, status, and statistics. All participants must be able to conduct their analyses in R so they can use the Data Explained platform.
- **If you would like to join this study as an analyst and collaborator, please do the following:**
    - **1) contact both Michael Feldman (feldman@ifi.uzh.ch) and Martin Schweinsberg (martin.schweinsberg@esmt.org) to receive your login details for the Data Explained platform**
    - **2) Visit www.dataexplained.org and complete a short survey about your personal and research background**
    - **3) conduct your analyses on the Data Explained website with the login details obtained in step 1. You will find the datafiles you'll need for your analysis in your Rstudio workspace after you log into your Data Explained account.**
    - **4) report your results using this short survey**

**Timeline:**

- Now until July 16th 2017: Analysis phase-- contributors analyze the dataset using the Data Explained platform
- July 16th 2017: Contributors submit their analytic approaches
- July 16th – September 1st 2017: Coordinators compile the results
- September 1st-September 31st 2017: Peer feedback and online discussion of the project results
- September 31st-December 31st 2017: Writeup of the final project report and comments from contributors via Google doc
- January-February 2018: Submission of the final report to a top academic journal for publication