

# Social corrections across party lines in a Twitter field experiment

Mohsen Mosleh\*<sup>1,2</sup>, Cameron Martel\*<sup>2</sup>, Dean Eckles<sup>2</sup>, David G. Rand<sup>2,3</sup>

<sup>1</sup>SITE Department, University of Exeter Business School; <sup>2</sup>Sloan School, MIT; <sup>3</sup>Department of Brain and Cognitive Sciences, MIT

ZOOM LINK:

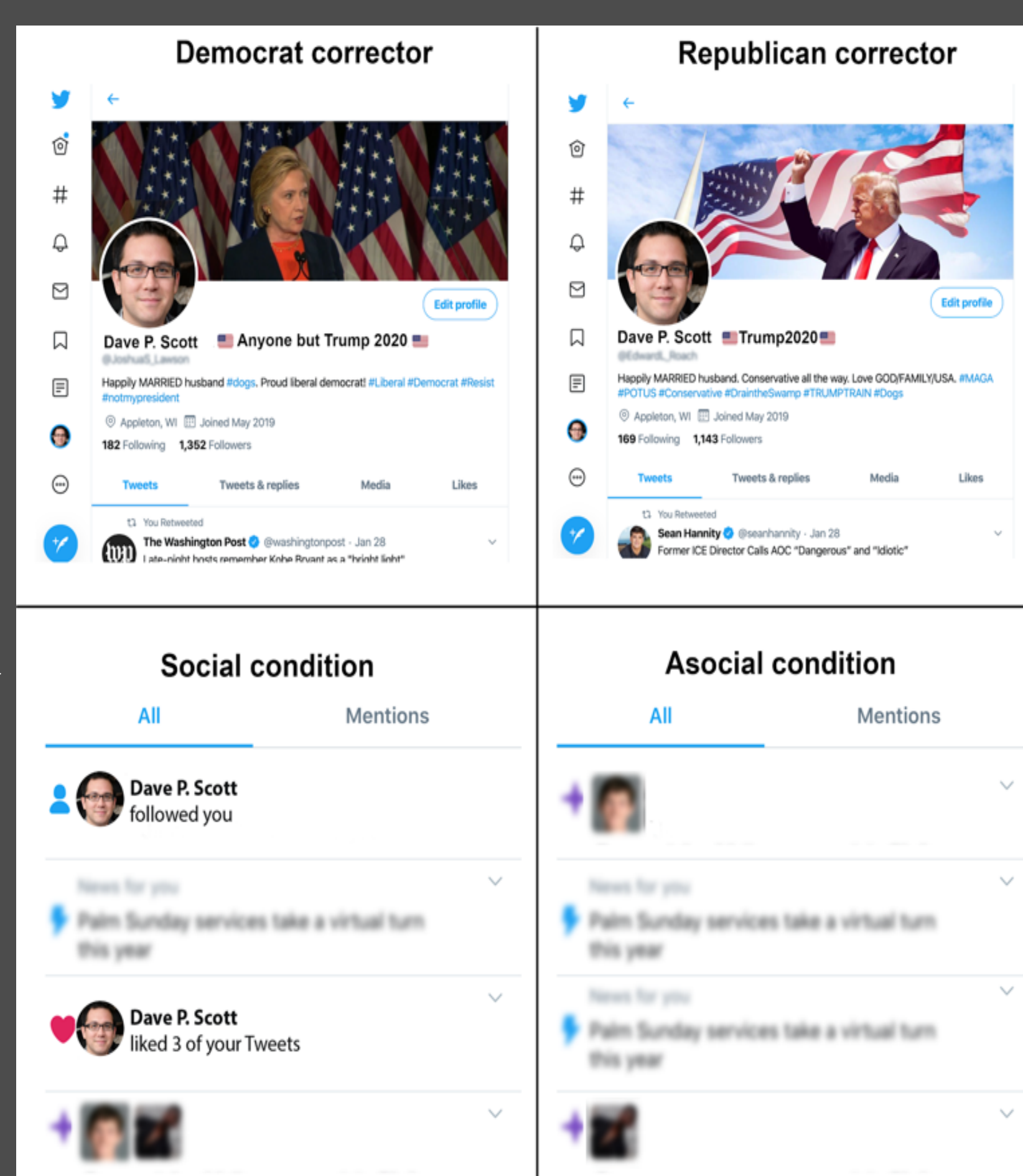
<https://yale.zoom.us/j/96006762641>

## Introduction

- **Misinformation** on social media is a major concern
- One way to combat falsehoods is through users correcting one another – ‘**social corrections**’ (1)
- We investigate what influences whether users ignore or engage with social corrections
- In particular, we examine the impact of:
  - Whether the corrected user has shared partisanship with the corrector
  - Whether the corrected user had **prior social contact** with the corrector

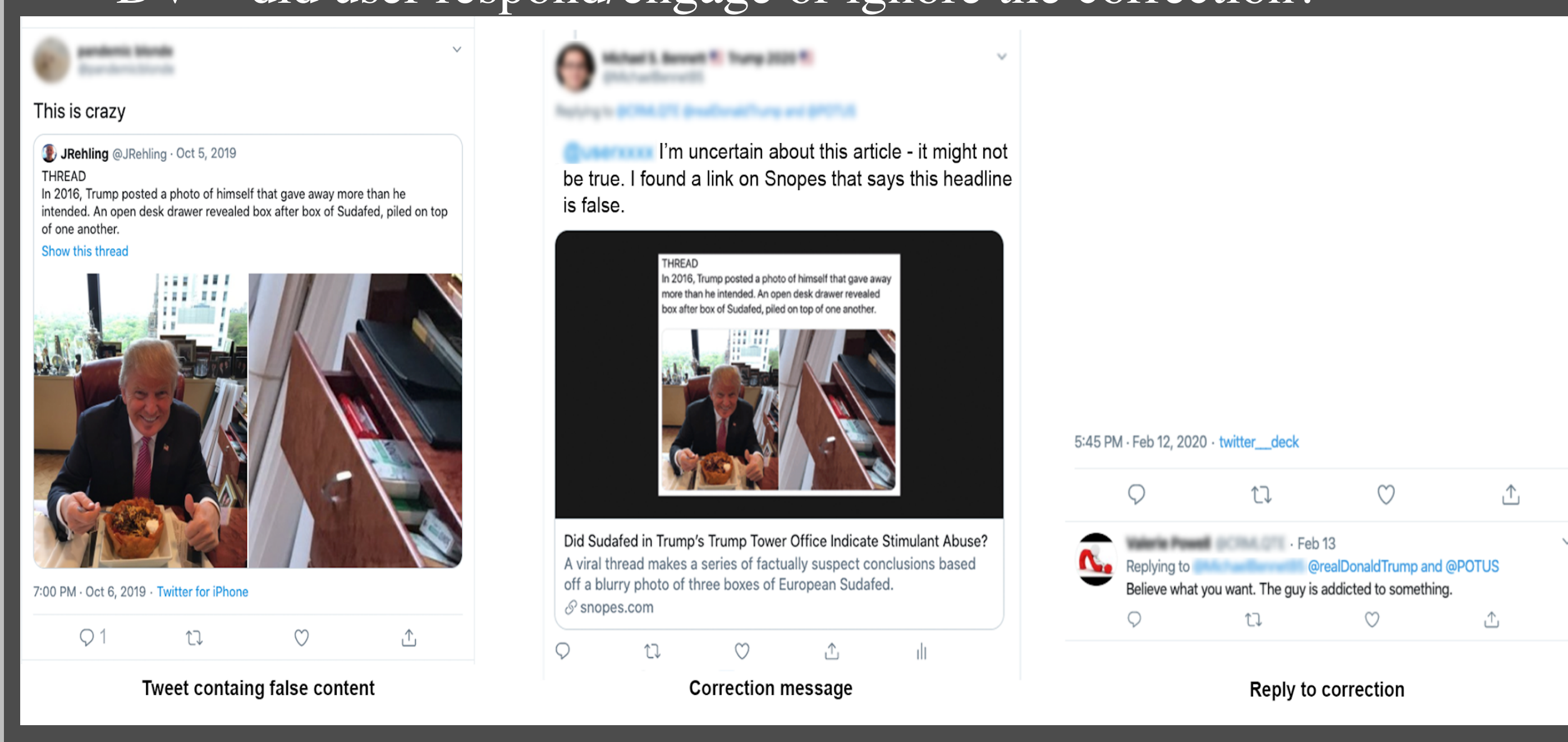
## Methods – Bots

- Retrieved 10 recent political stories rated false by Snopes.com; identified **1,454 Twitter users** who shared one of these stories
- Constructed a set of bots: **half Democrat, half Republican**. Also varied prior social interaction with soon to be corrected users – **half of bots followed users ahead of time & liked several of their tweets, and half did not.**

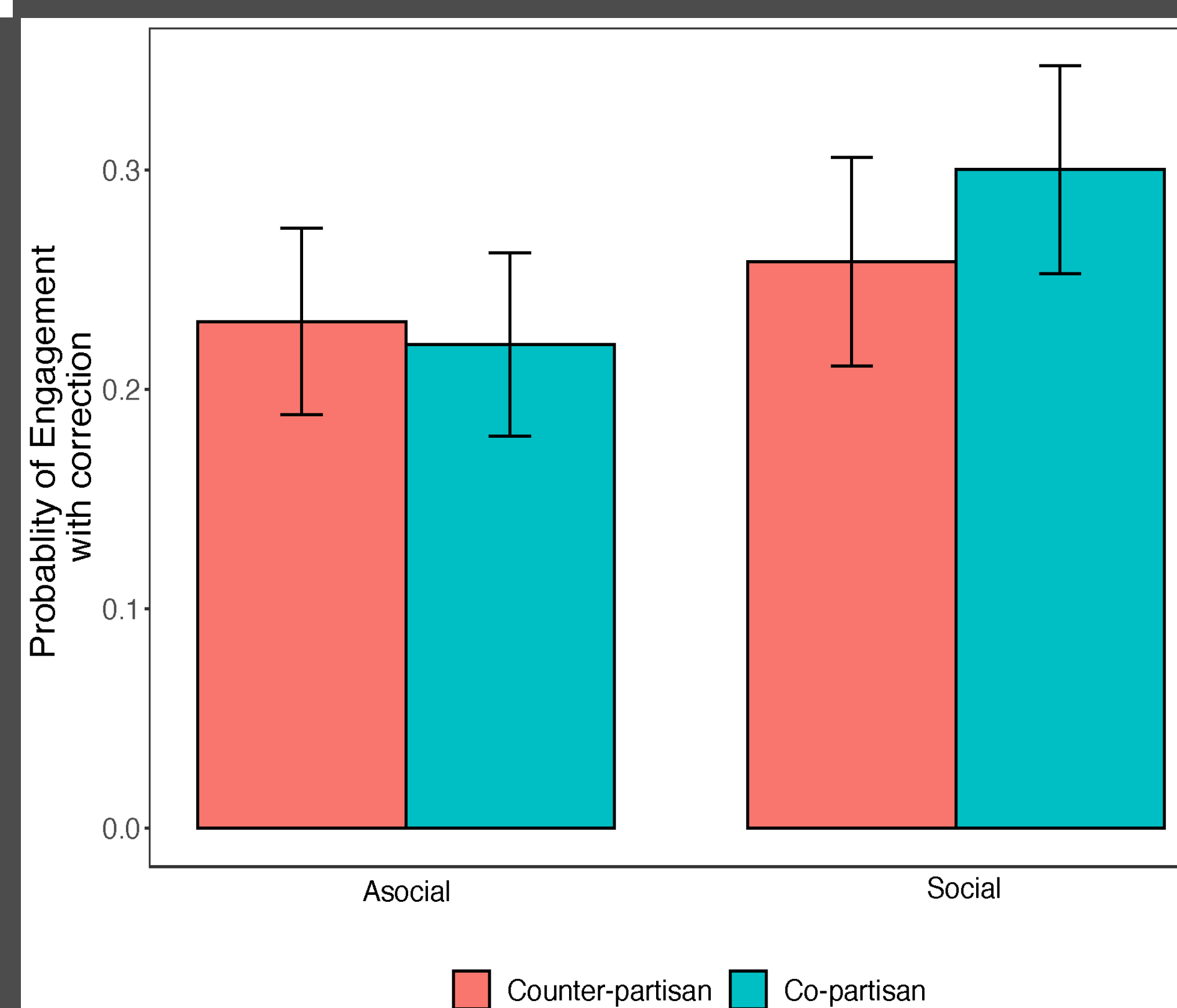


## Methods – Correction Messages

- For each user, a bot issued a correction message indicating that article is false, and providing a Snopes link.
- DV = did user respond/engage or ignore the correction?



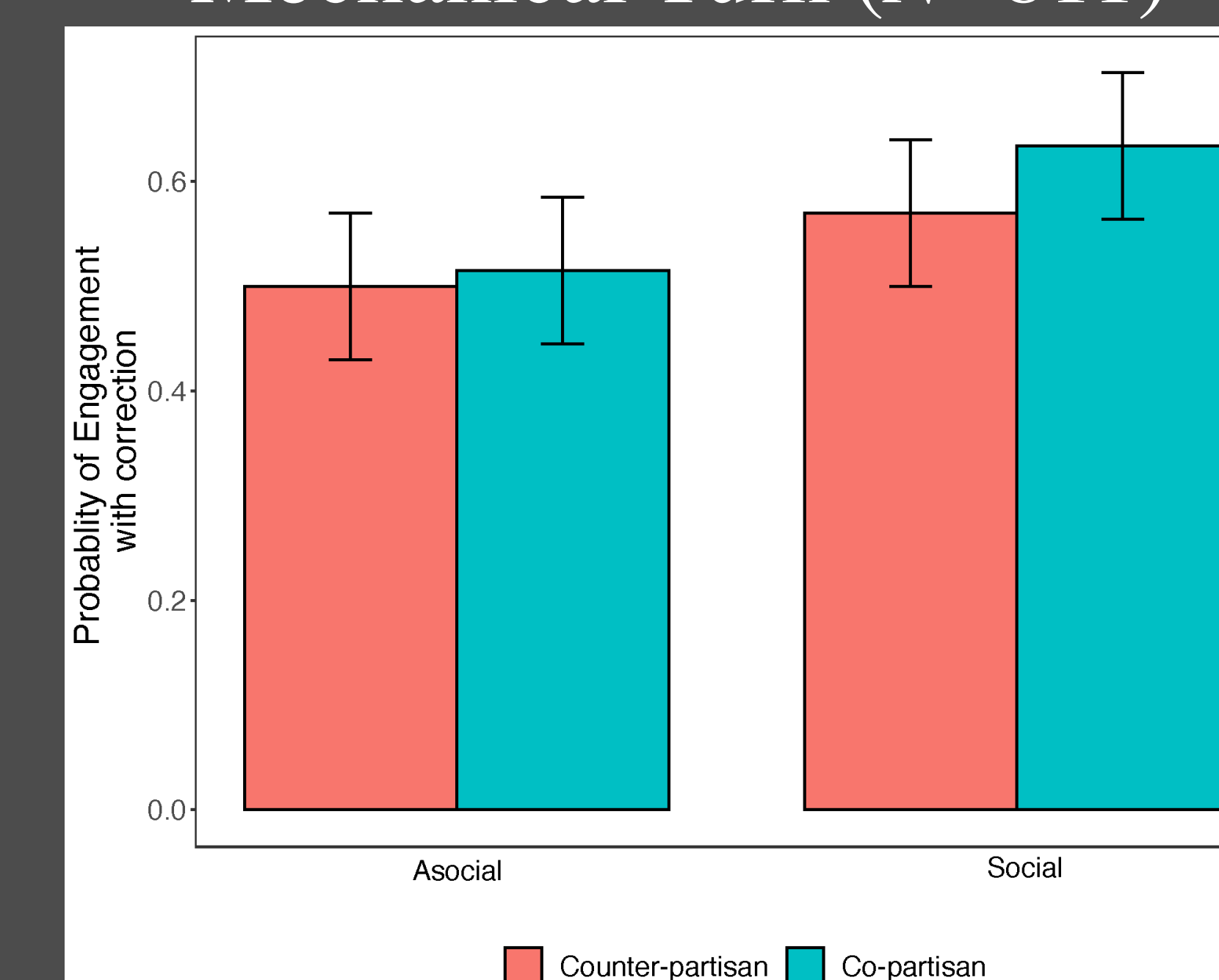
## Results



- We used linear regression and Fisherian Randomization predicting **engagement** (0 or 1) given **partisan alignment** (-0.5=Different Party, 0.5=Same Party) and **prior social interaction** (-0.5=Asocial, 0.5= Social)
- **Significant effect of prior social interaction** ( $b=0.053, p_{reg}=0.018, p_{FIR}=0.016$ )
- No evidence of effect of partisan alignment
- No interaction between prior social interaction and partisan alignment
- No significant difference in effect of positive vs negative responses

## Study Experiment Replication

Replication of our Twitter field experiment using self-report intentions on Amazon Mechanical Turk (N=811)



- **Replicated significant effect of prior social interaction**

## Discussion

- **Minimal social connections** on social media may **increase efficacy of corrective messages**
- **No evidence** that people are more likely to ignore corrections from **counter-partisans**
- Results shown both in a **Twitter field experiment** and an **MTurk survey experiment**

## References

1. Vraga, E. K., & Bode, L. (2018). I do not believe you: how providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21, 1337-1353.