

Pooling Resources - A Comparison Study of Online Subject Pools and their Suitability for Various Psychology Experiments



Constantin Rezlescu, Sidharth Prabhu-Naik

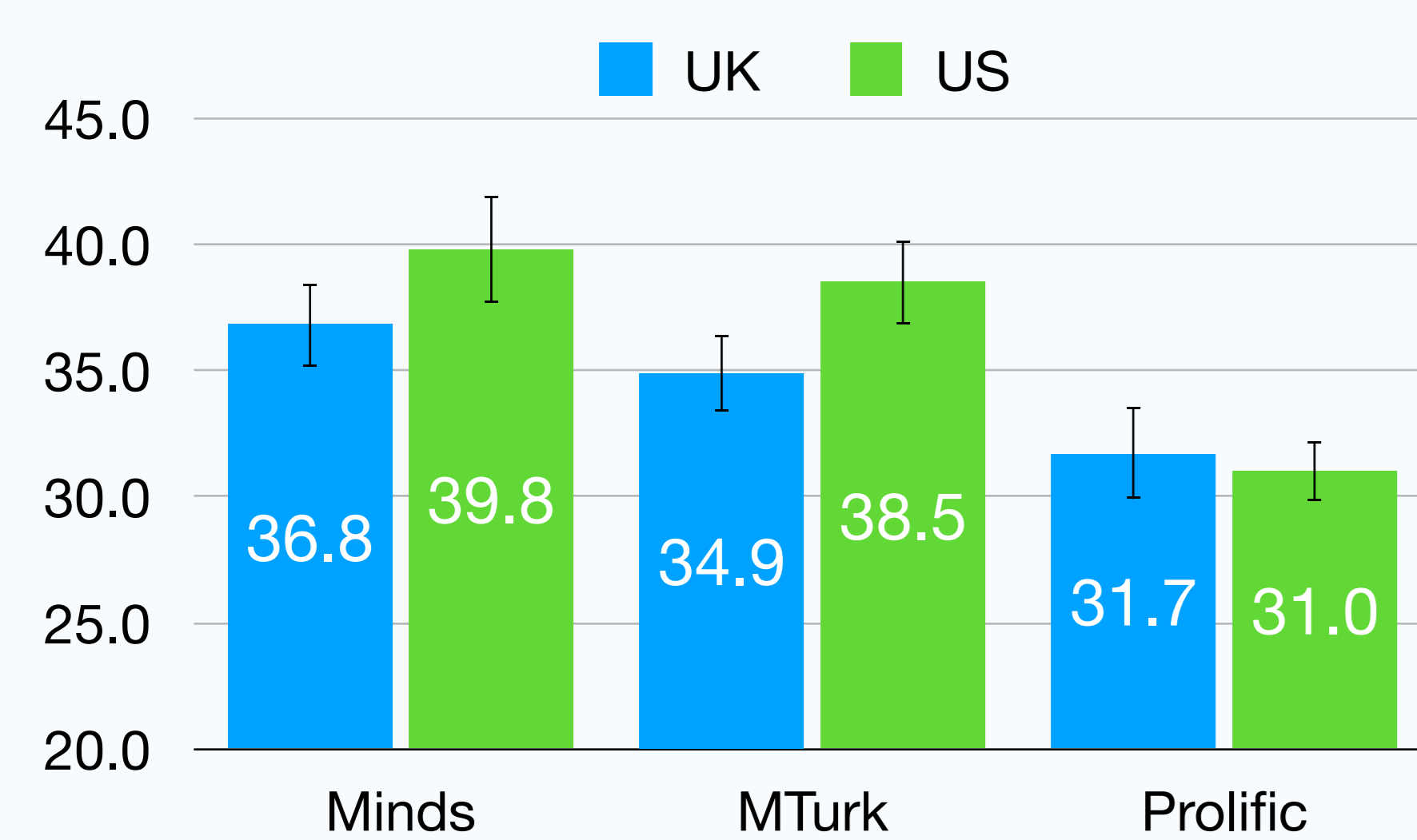
Department of Psychology, University College London

Introduction

- Not long ago, experimental psychologists relied almost exclusively on laboratory settings and restricted participant pools (i.e. undergraduate students) for their data
- More recently, an increasing number of researchers conduct their studies online (Zhou et al., 2016), recruiting more diverse participants quickly and at a lower cost from services such as Amazon Mechanical Turk (MTurk)
- However, there are **reservations about the reliability and quality** of the data generated by online participants (e.g. Ford, 2017), mostly regarding the **veracity of the self-reported demographics** (e.g. native language), **capacity to follow instructions, attention and effort during experiments**
- **Current study aims to investigate if an advanced participant verification system, involving ID check and face authentication, can produce better data and lower rates of exclusions with online participants**
- Secondary aim was to investigate possible **differences in personal characteristics** between participants enrolled in different subject pools

Methods

- We recruited 100 participants (50 US-based and 50 UK-based) from each of these subject pools: **MTurk, Prolific, Testable Minds**
- Testable Minds offers the possibility to recruit verified participants, who undergo an **advanced ID verification process and face authentication** (see Rezlescu et al., 2020, for details)
- Total participants: **N = 300** (147 female, 151 male, 2 other)
- Mean age: 35.5 years old (SD = 12.2)



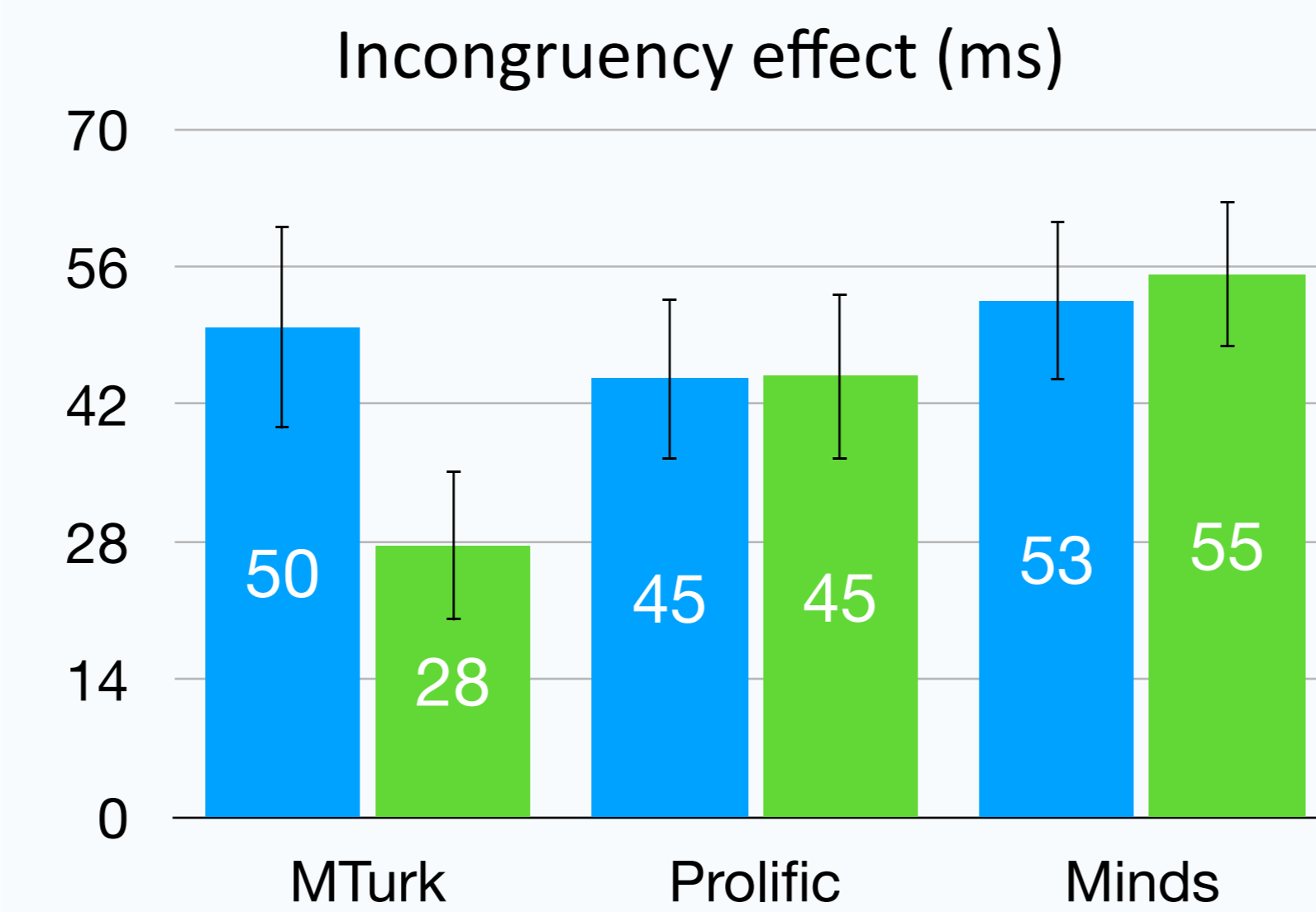
Prolific participants were significantly younger than participants from MTurk and Testable Minds

- Participants were asked to complete:

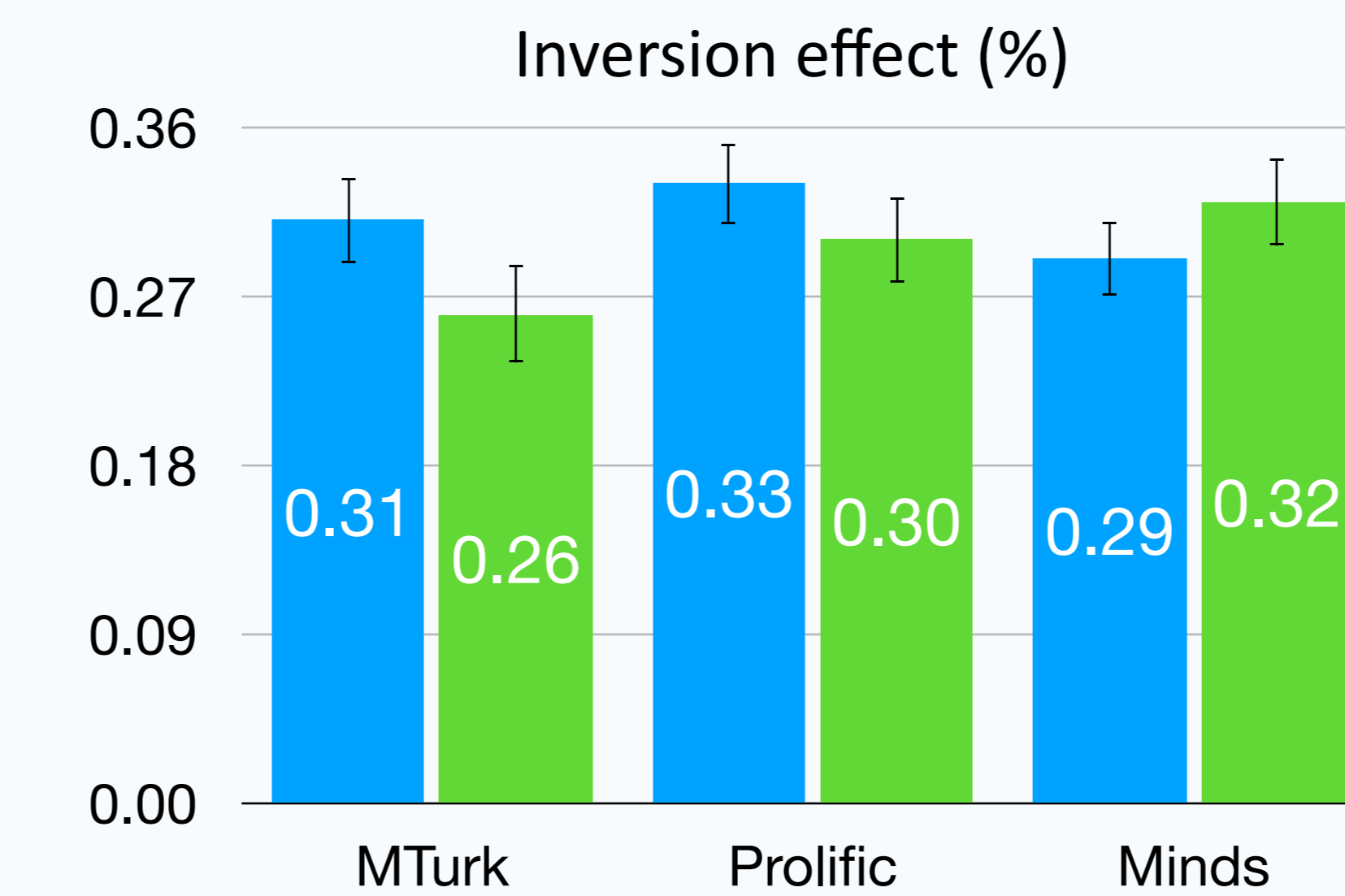
- **seven classic psychology tests:** Flanker, Face inversion, Serial position, Semantic priming, False memory, Anchoring, Asian Disease (not reported here)
- the **Ten Item Personality Inventory** (Gosling et al., 2003)
- the **Depression, Anxiety and Stress Scale** (DASS-21; Henry & Crawford, 2005)
- one **instruction manipulation check** (a “catch” question) to detect participants who do not read or comprehend instructions

Classic psychology effects

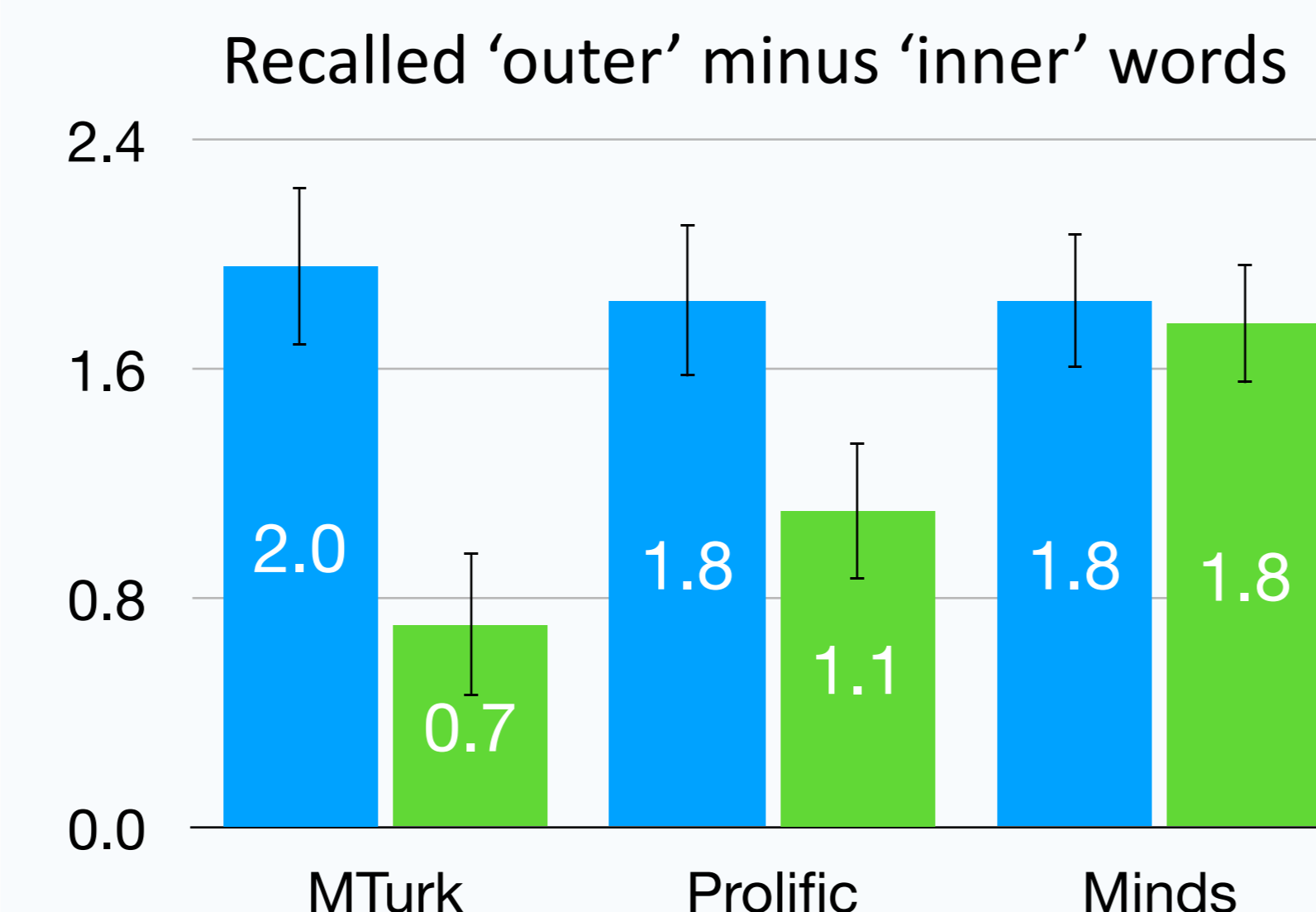
Flanker task



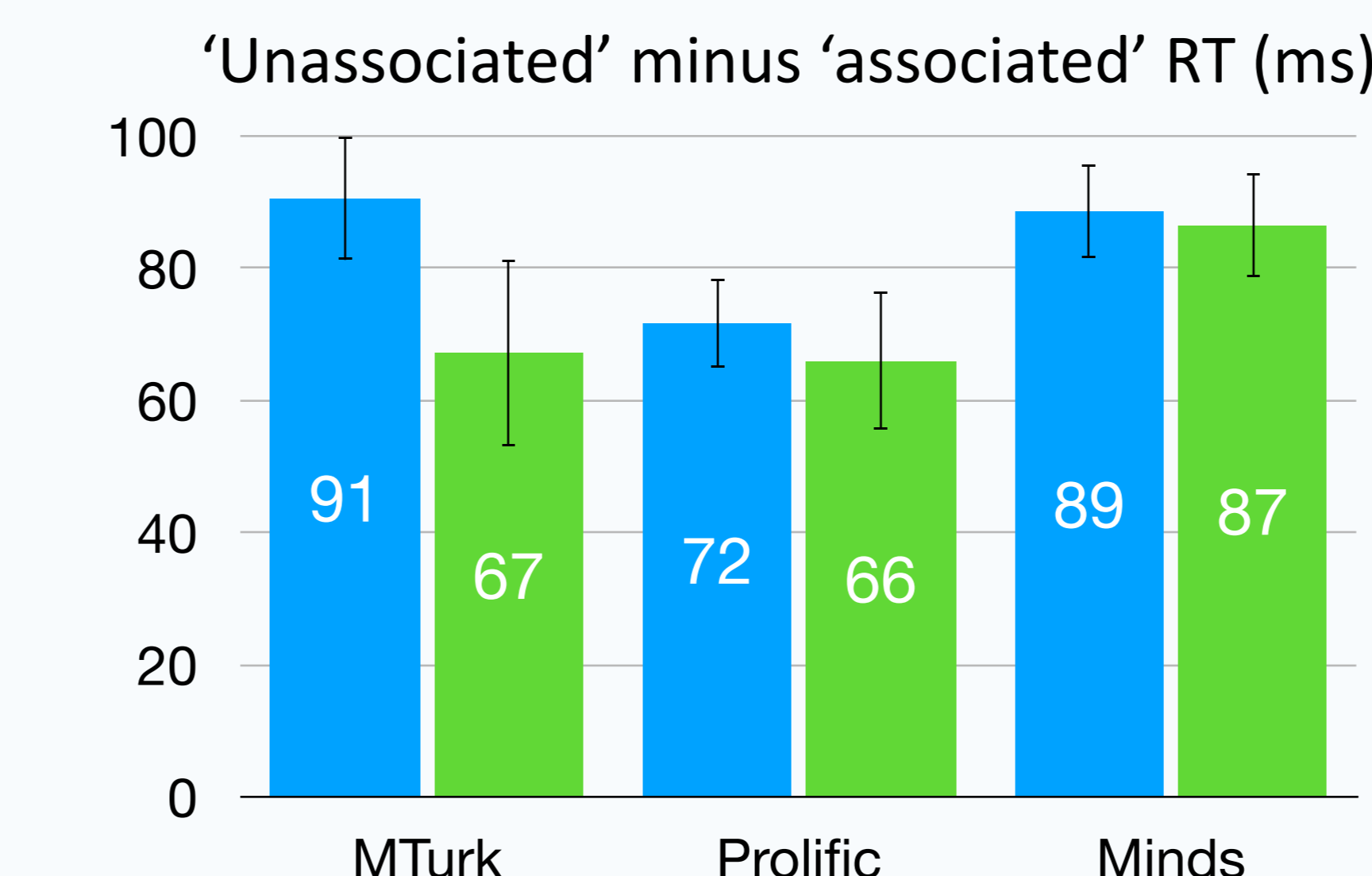
Face inversion



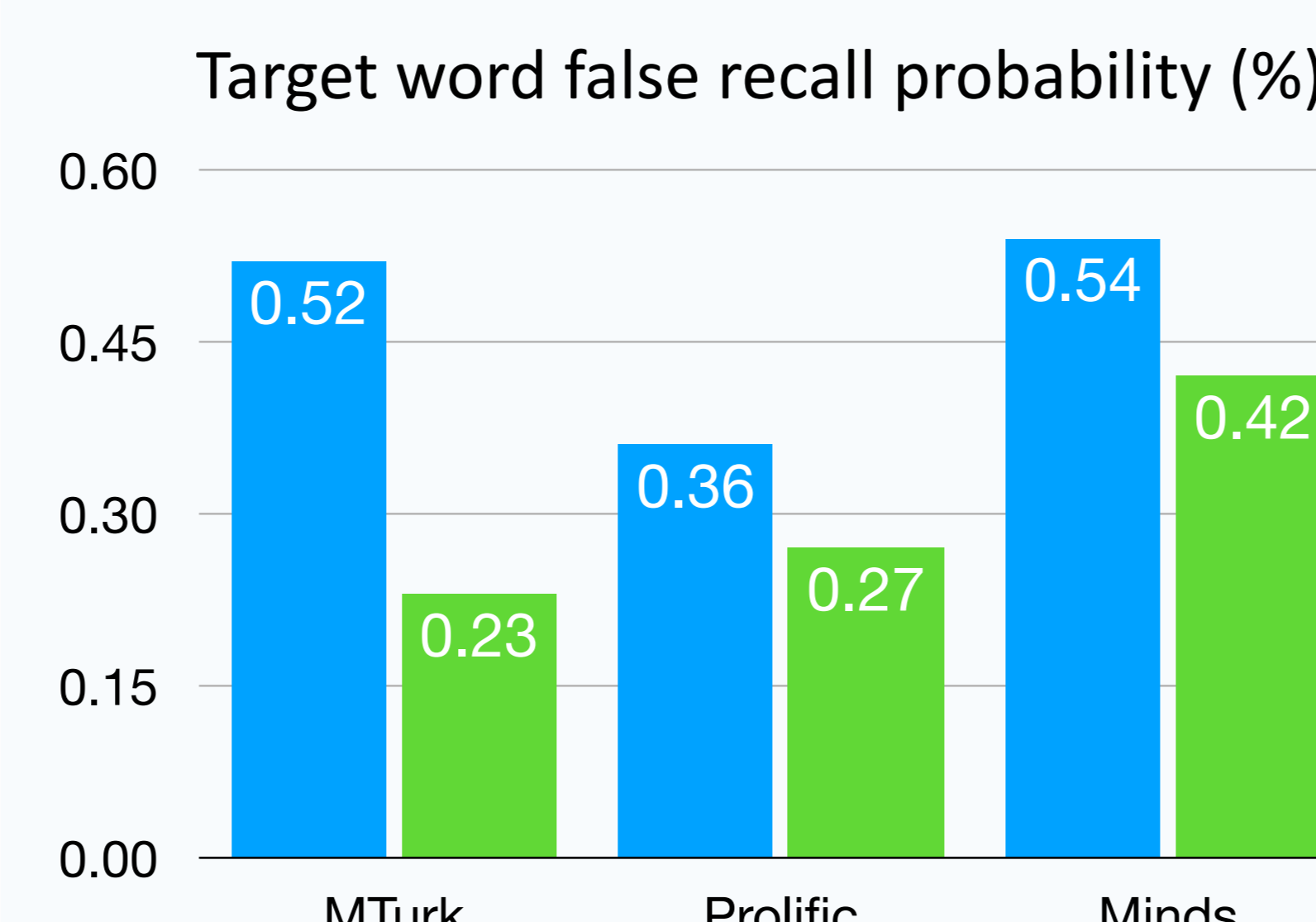
Serial position



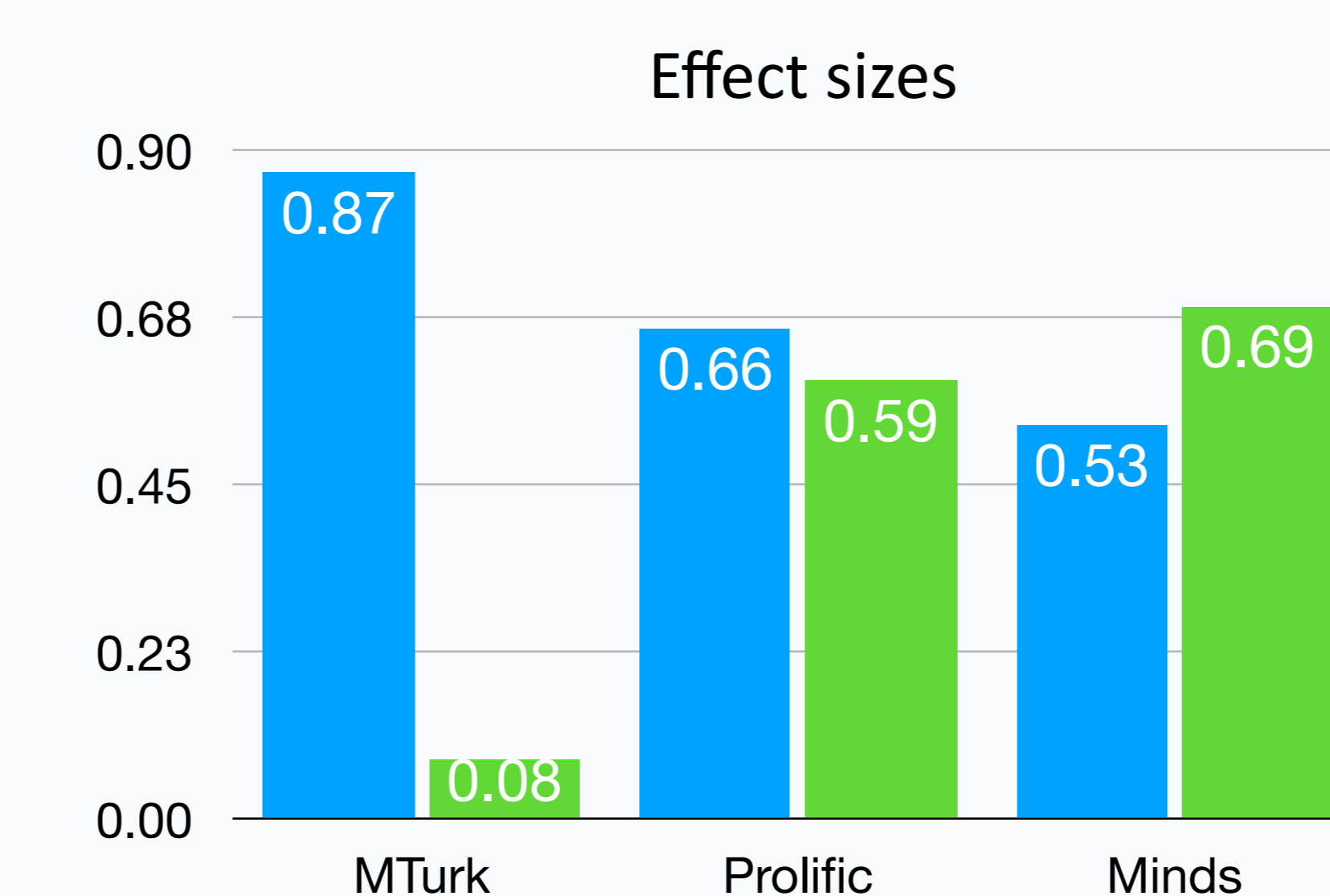
Semantic priming



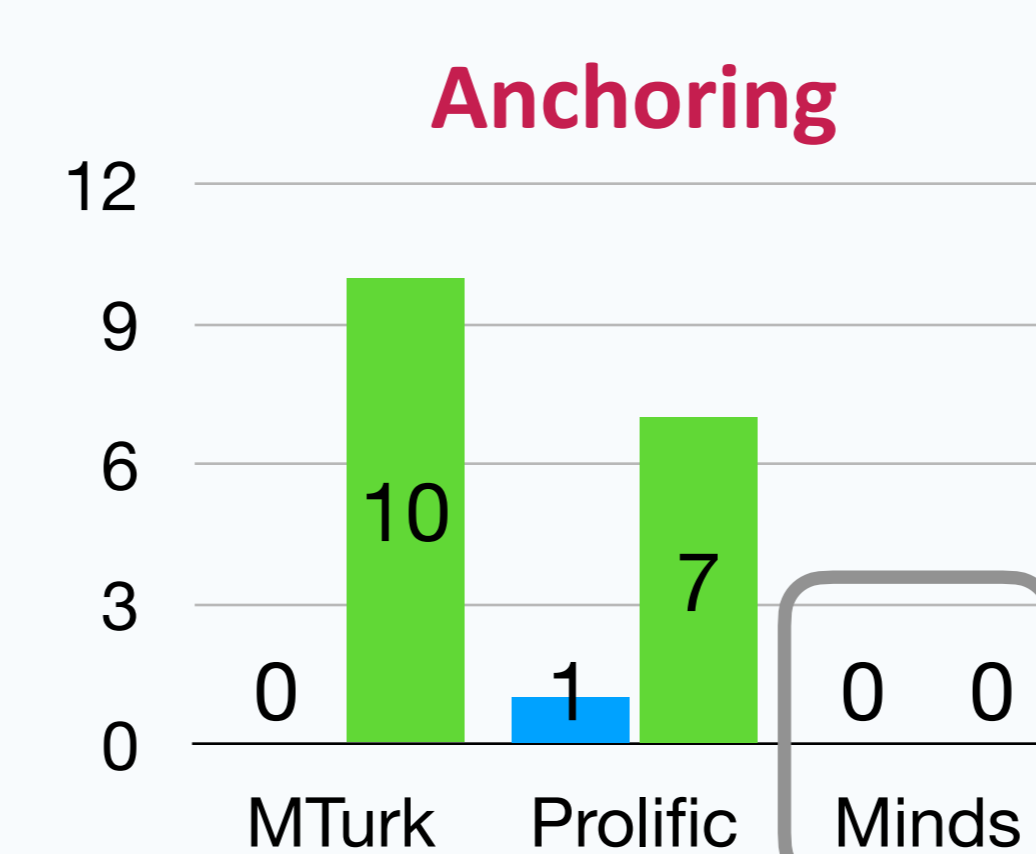
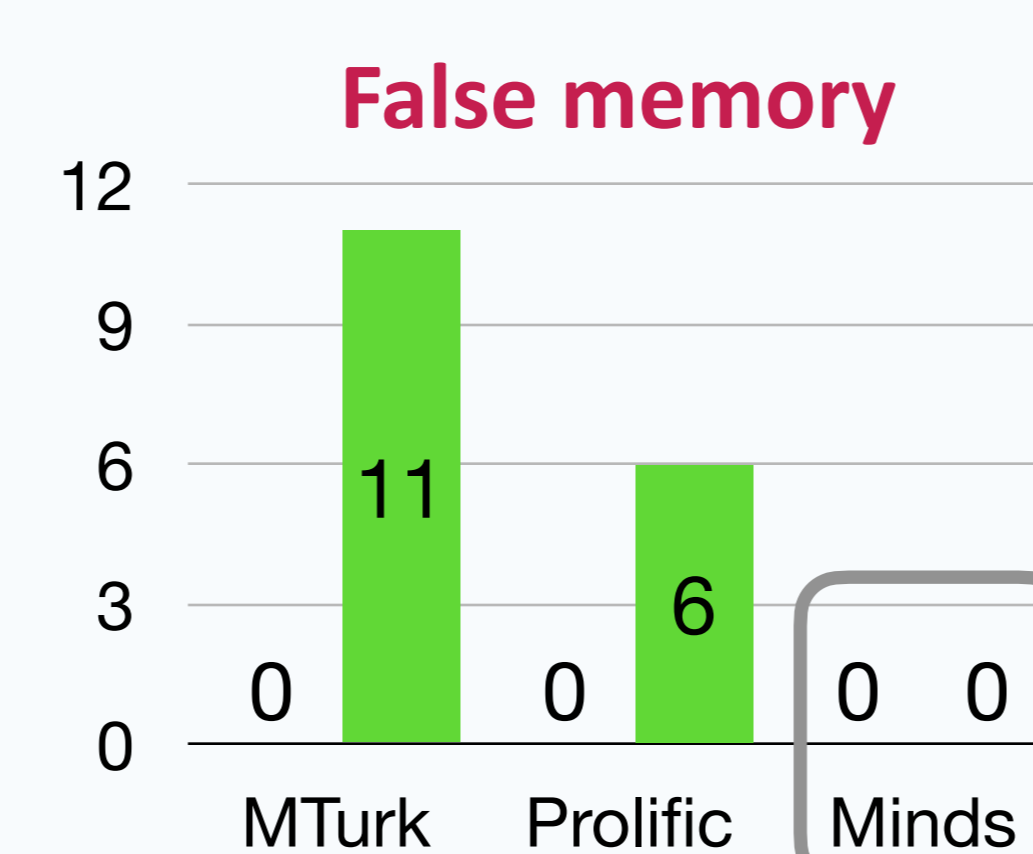
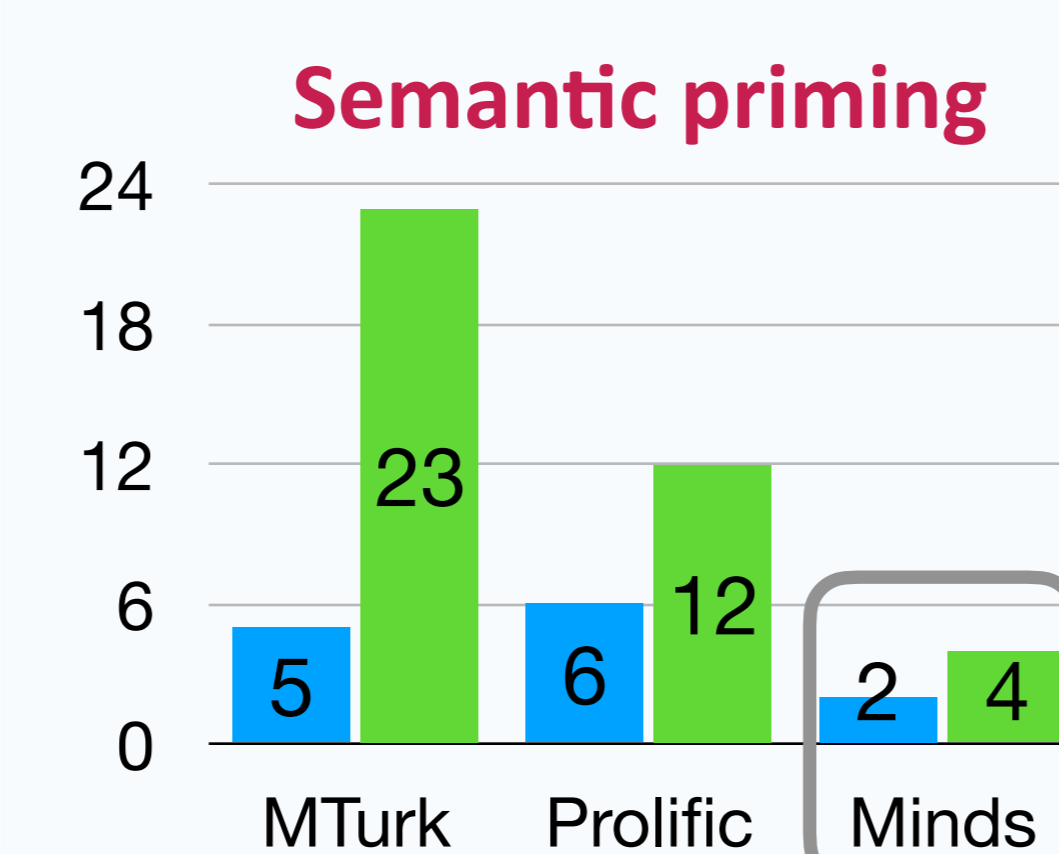
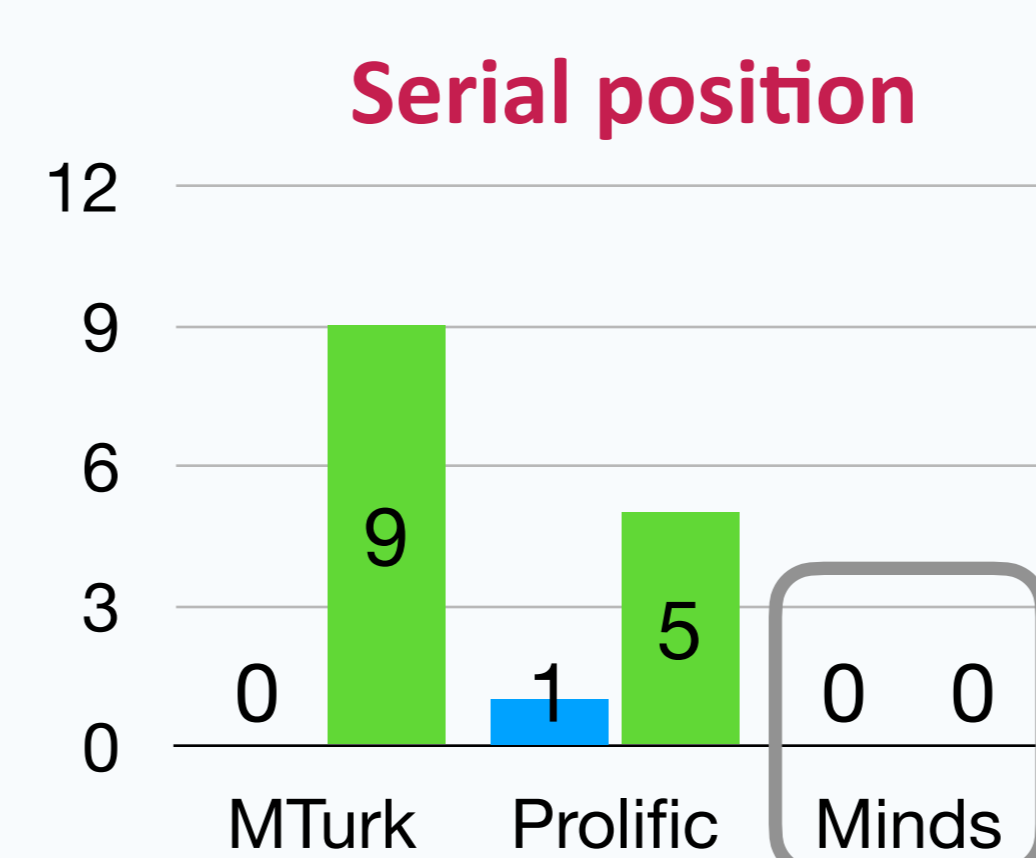
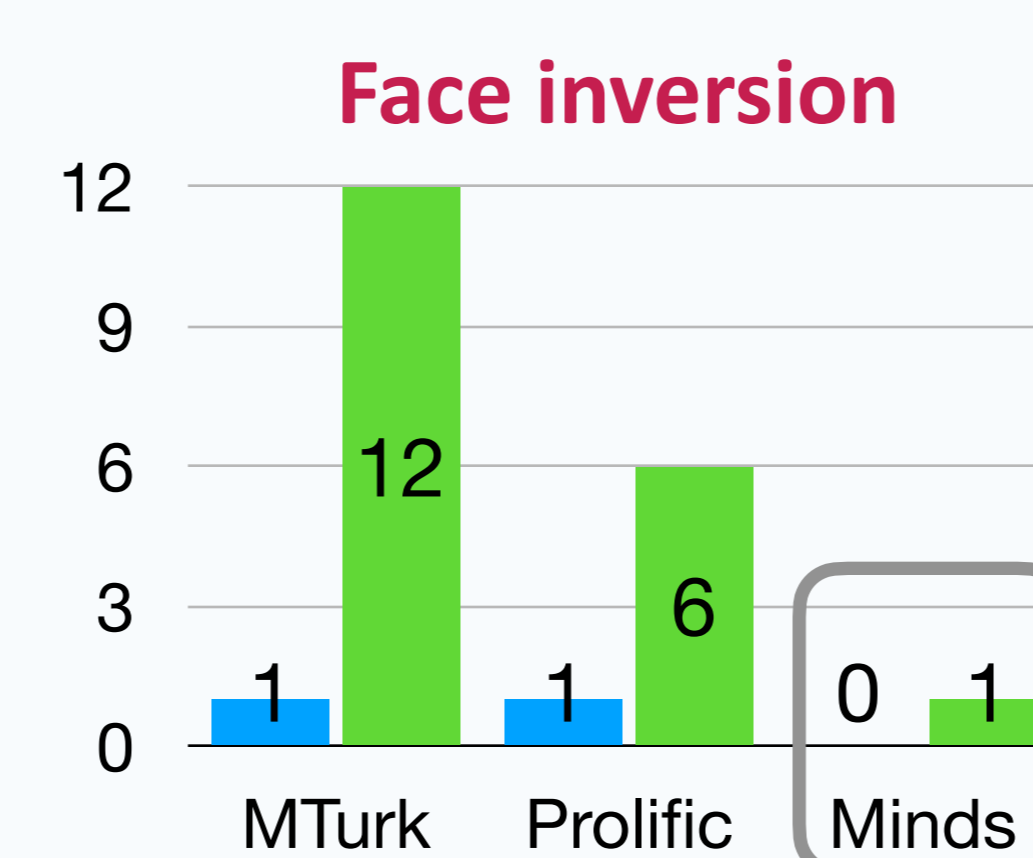
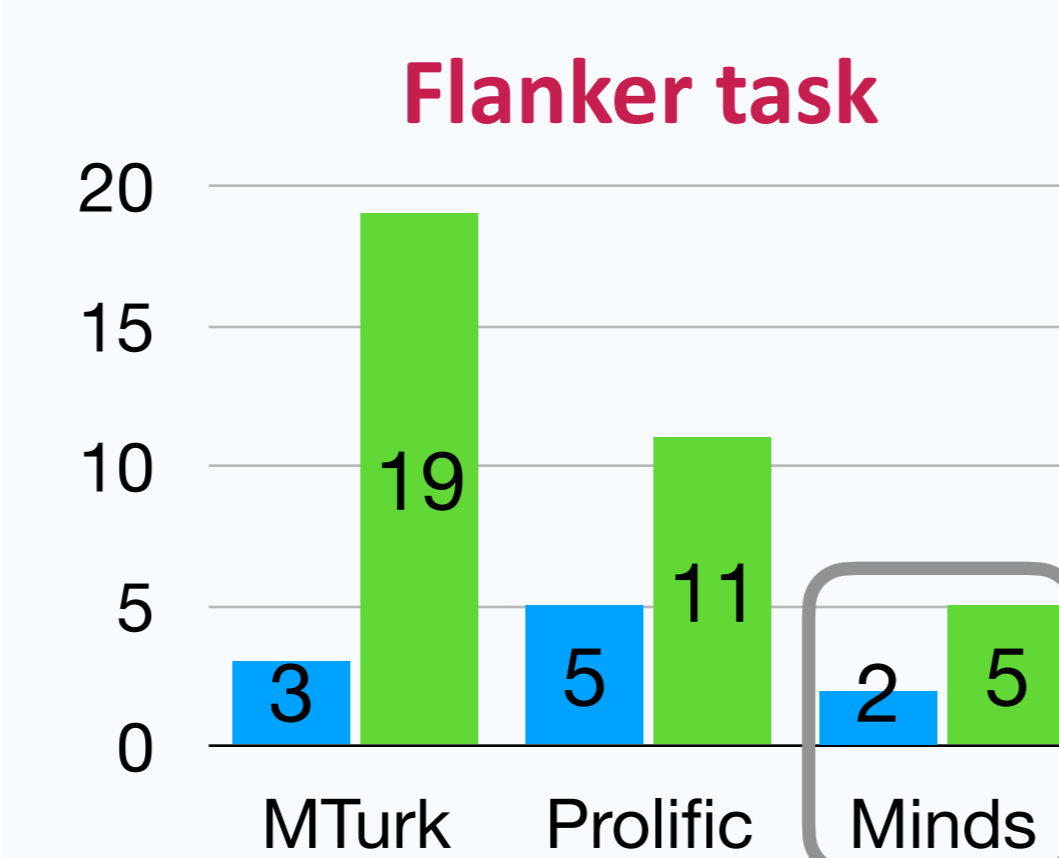
False memory



Anchoring

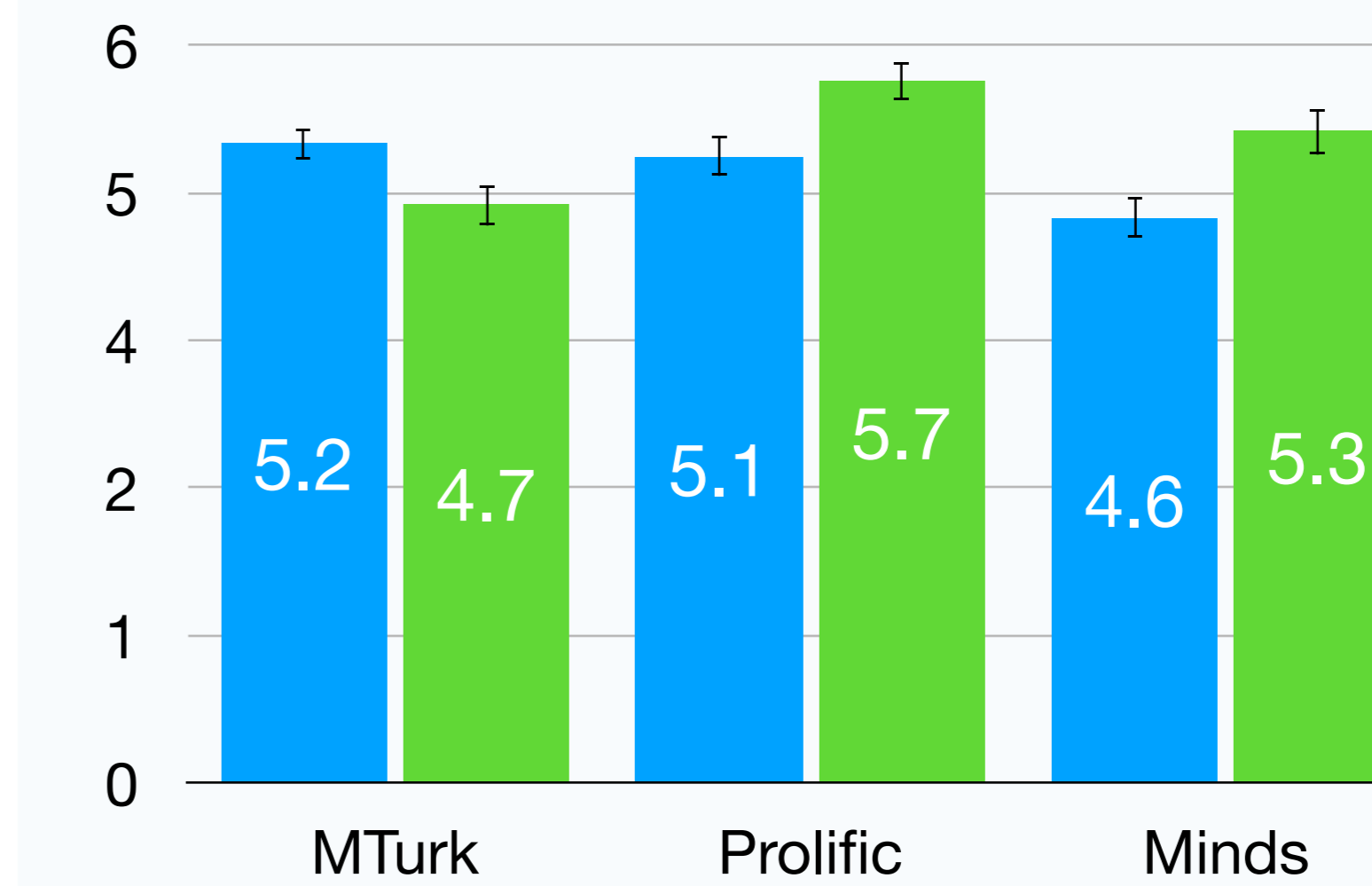


Number of participants excluded (specific criteria set per task)

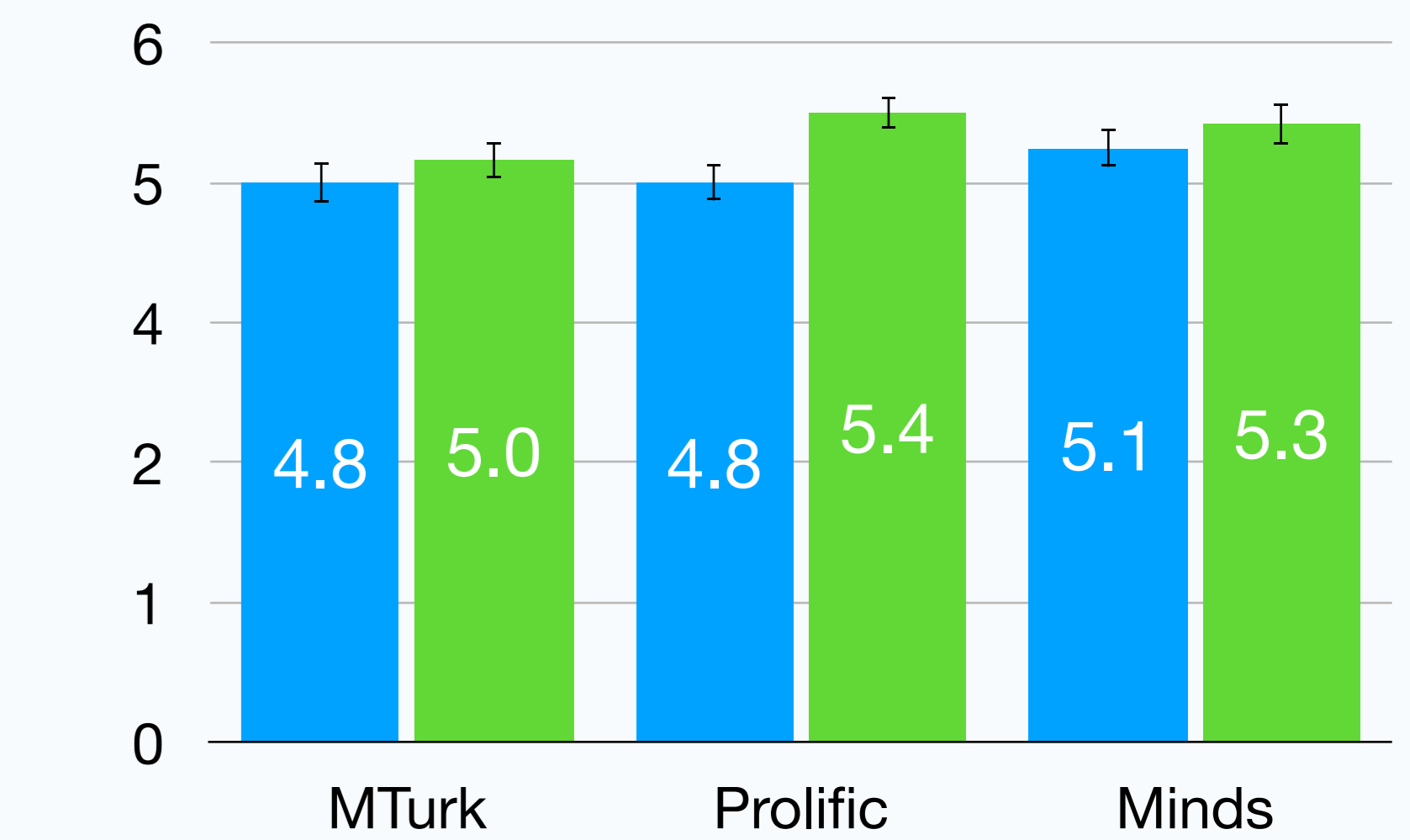


Personal characteristics

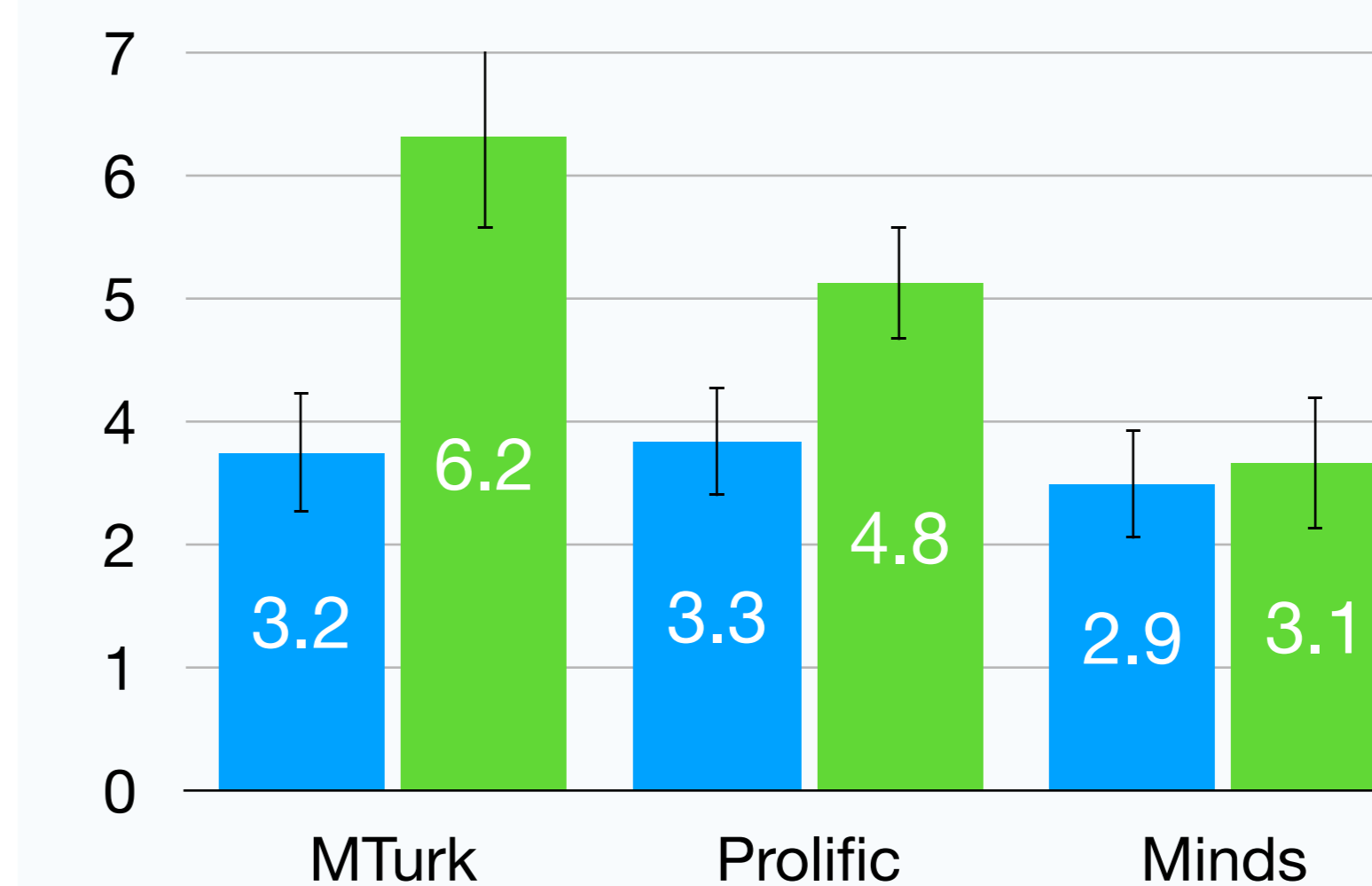
Openness



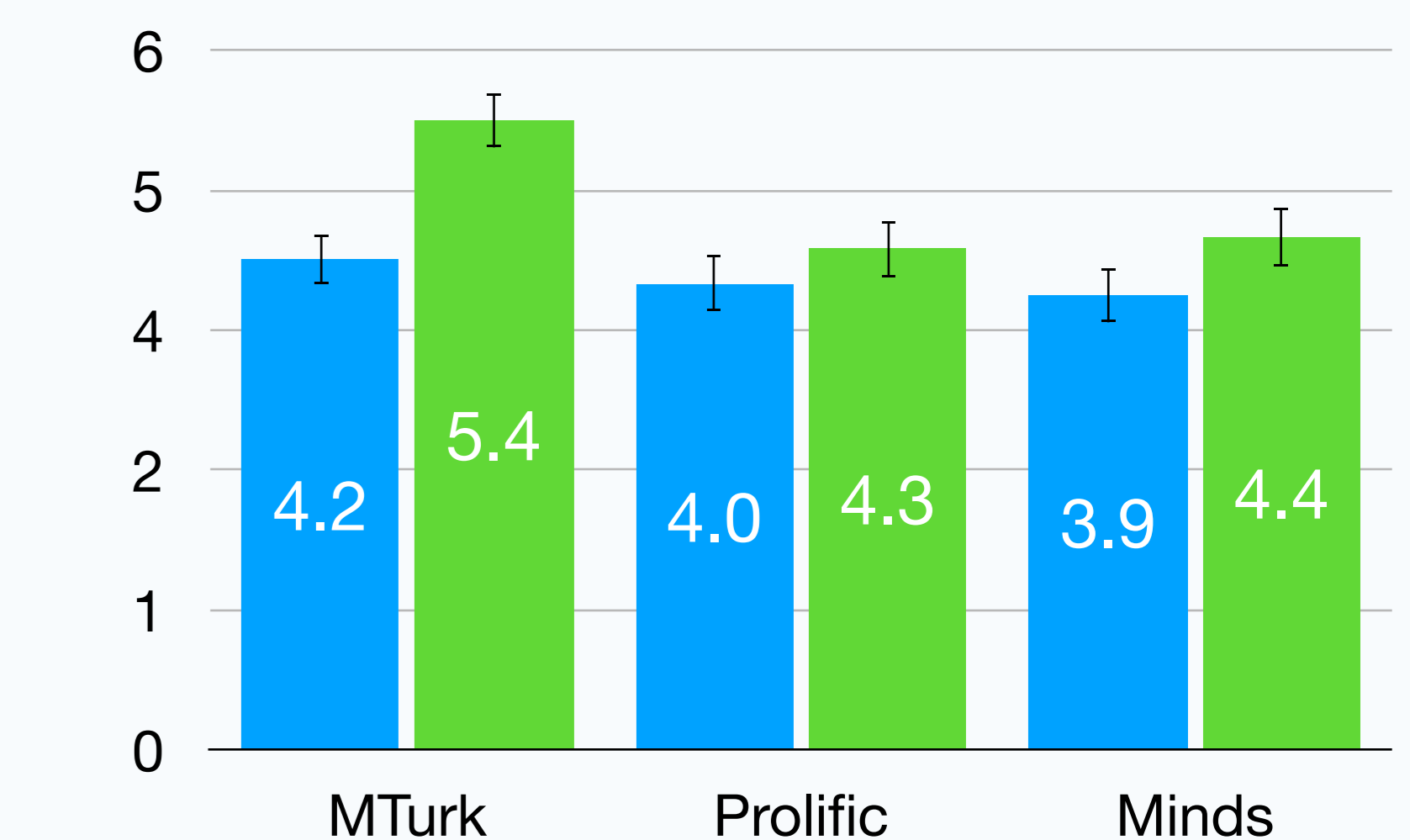
Agreeableness



Anxiety

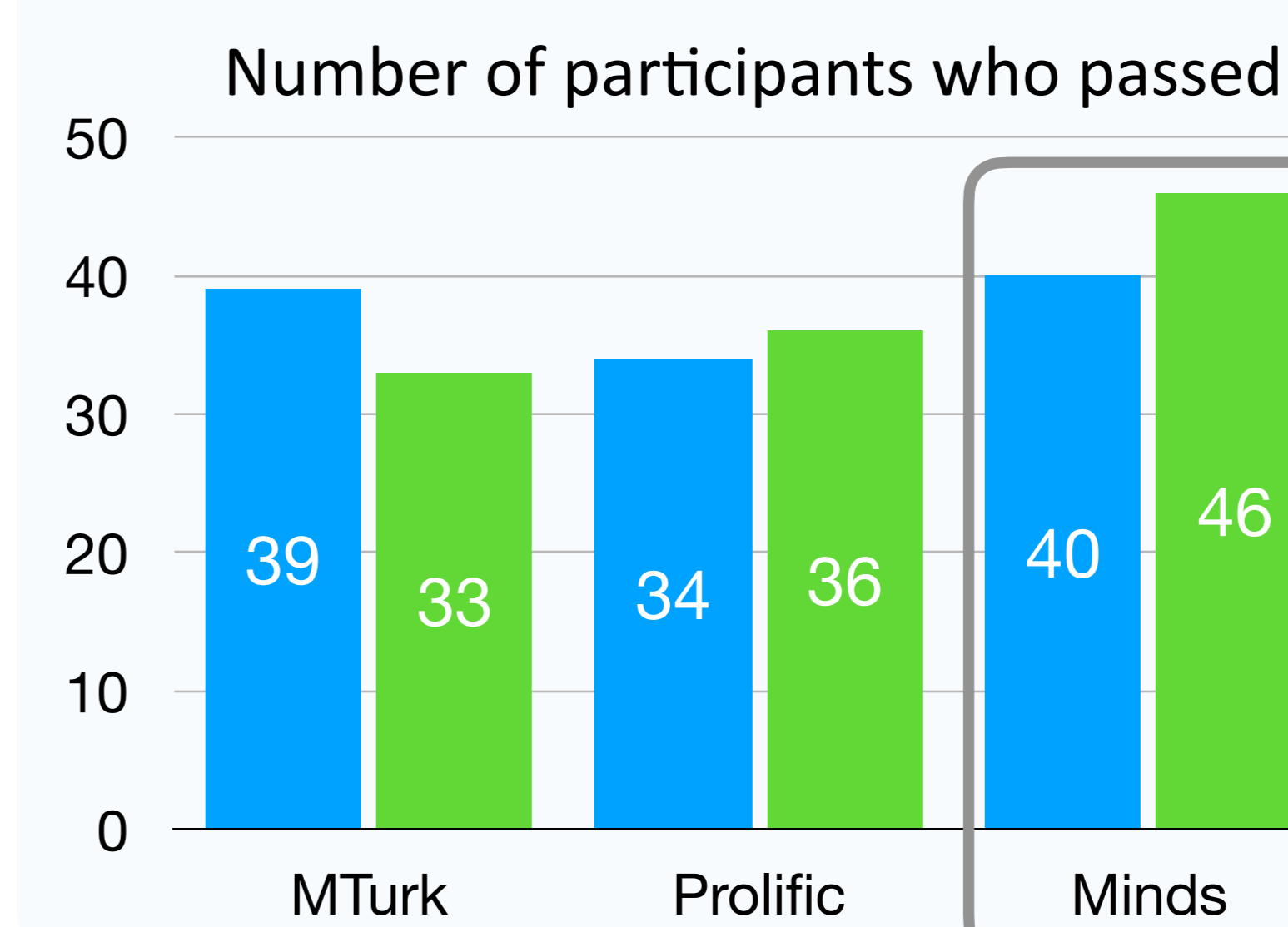


Self-esteem

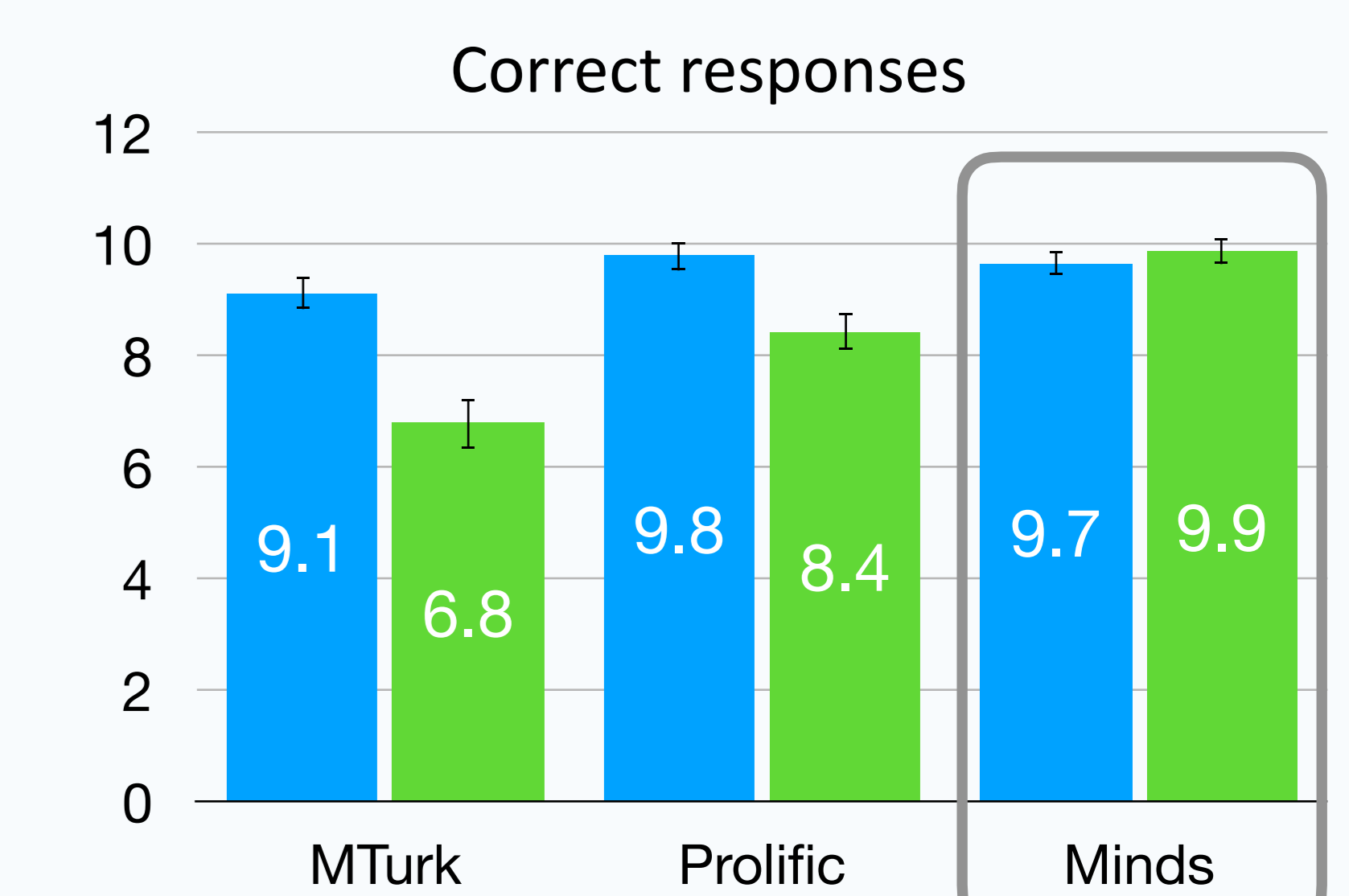


Understanding instructions

Instructional manipulation check



English knowledge test



Conclusions

- Verified online participants from Testable Minds had the lowest number of exclusions in each experimental task, largest number to pass the instructional manipulation check, and the highest scores for the English test
- Overall, participants from Testable Minds also tended to show the largest effects in the experimental tasks (though not always), similar to the effects observed in the lab
- This suggests that **an advanced ID verification and face authentication system have the potential to lead to better quality for online data, similar to that collected in the lab**
- MTurks seemed to be the least reliable, with the US sample particularly problematic
- There were some significant differences in terms of personal characteristics for participants from different pools, which may be important for certain studies