How do researchers interpret the results of multiple experiments?

Olmo van den Akker, 14-03-2019

Introduction

"It is as necessary to be able to [do statistics] as it is to read and write"

Wells

rpret

t the

Experiments 4, 5, and 6: The Experimental-Causal-Chain Approach

Because all variables were measured in Experiment 3, it is not possible to establish causality in our mediation, that is, we cannot definitively say that power drove the reported differences in unethical behavior. To address this inherent limitation and provide complementary evidence for the underlying role of power, Experiments 4, 5, and 6 used the experimental-causal-chain approach (Spencer et al., 2005). Experiment 4 manipulated social class and measured how it affects people's sense of power. Experiments 5 and 6 manipulated power and examined its effects on unethical behavior that either benefited the self or others. If differences in power underlie our effects, manipulating power should lead to the same behavioral tendencies as measured by differences in social class.

Experiment 4: Manipulating Social Class

Experiment 4 manipulated social class and measured its impact on sense of power.

Method. One hundred fifty participants (79 female; M_{age} = 21.15, SD = 2.82) were randomly assigned to a three-cells design (status: high-social-class vs. low-social-class vs. baseline). Participants, mostly students from a large European metropolitan area,

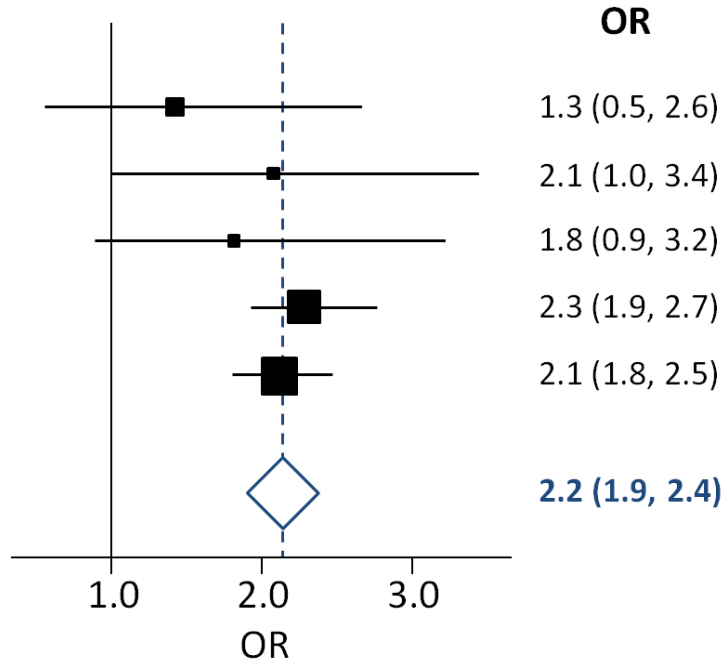
Experiment 5: Manipulating Power

In Experiment 5, we directly manipulated power and expected that high power would increase self-beneficial unethical behavior, but low power would increase other-beneficial unethical behavior, similar to the behavioral tendencies triggered by differences in social class (Experiments 1 through 3).

Method. One hundred twenty-one participants (59 females; M_{age} = 20.56, SD = 1.17) were randomly assigned to a 3 (power: high vs. low vs. baseline) × 2 (lie beneficiary: self vs. other) between-subjects design. Participants were students of a large Midwestern University participating in a lab study in exchange for $12.

Power manipulation. Participants completed two independent tasks in the lab. In the first task, participants completed a power recall task (Galinsky et al., 2003) that has been widely used to elicit differences in feelings of power (for reviews, see Galinsky et al., 2015; Rucker et al., 2012). In the high-power condition, participants wrote about a time they had power. In the low-power condition, participants wrote about a time they lacked power. In the baseline condition, participants wrote about the last time they went to the grocery store.

Unethical behavior measure. Participants then completed a second task, portrayed as a decision-making task for the psychology department. They were presented with a series of three scenarios in



TILBURG UNIVERSITY

People

Experiment:

Follow-up analysis:

Supervision:

The Sample

Participants: researchers, editors, and reviewers in two fields of psychology

Social psychology (1126 papers):

- JPSP
- JESP
- PSPB
- EJSP

Experimental psychology (1323 papers):

- JEP: G
- JEP: HPP
- JEP: LMC
- QJEP
- C&E

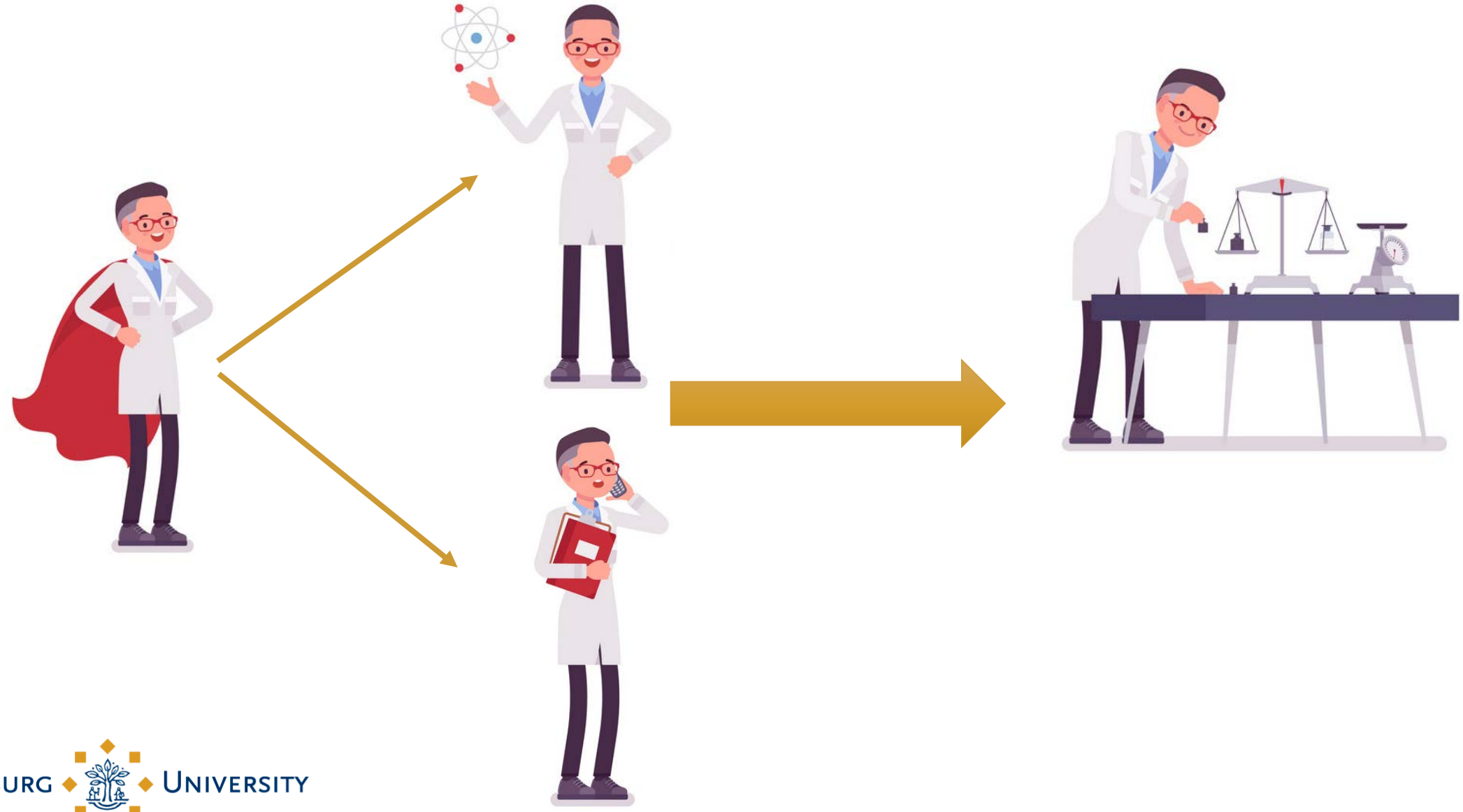
Contacted:

1810 authors
834 editors/reviewers

Participants:

505

The Experiment

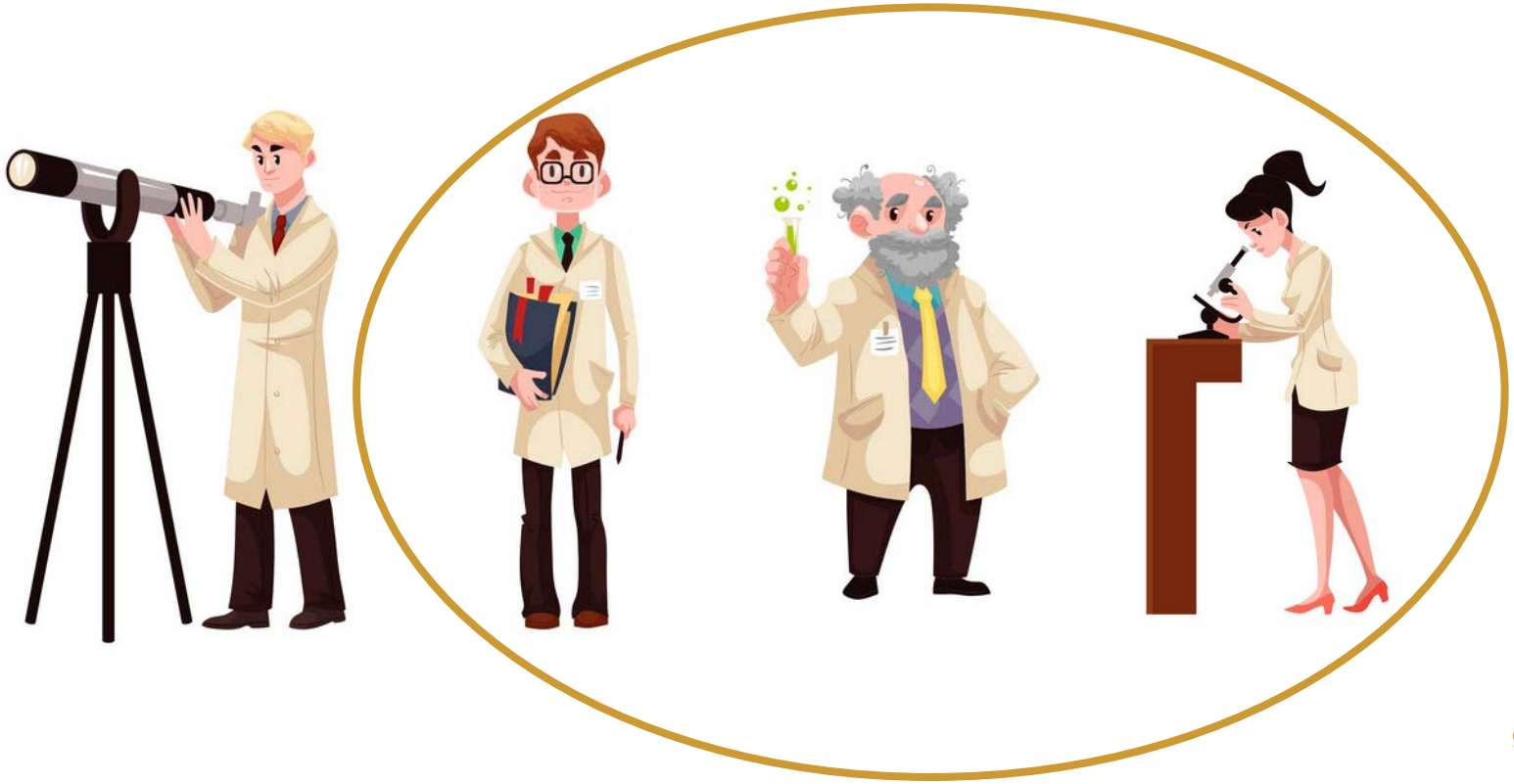


The Experiment

The Experiment

The Experiment

The Experiment



The Experiment

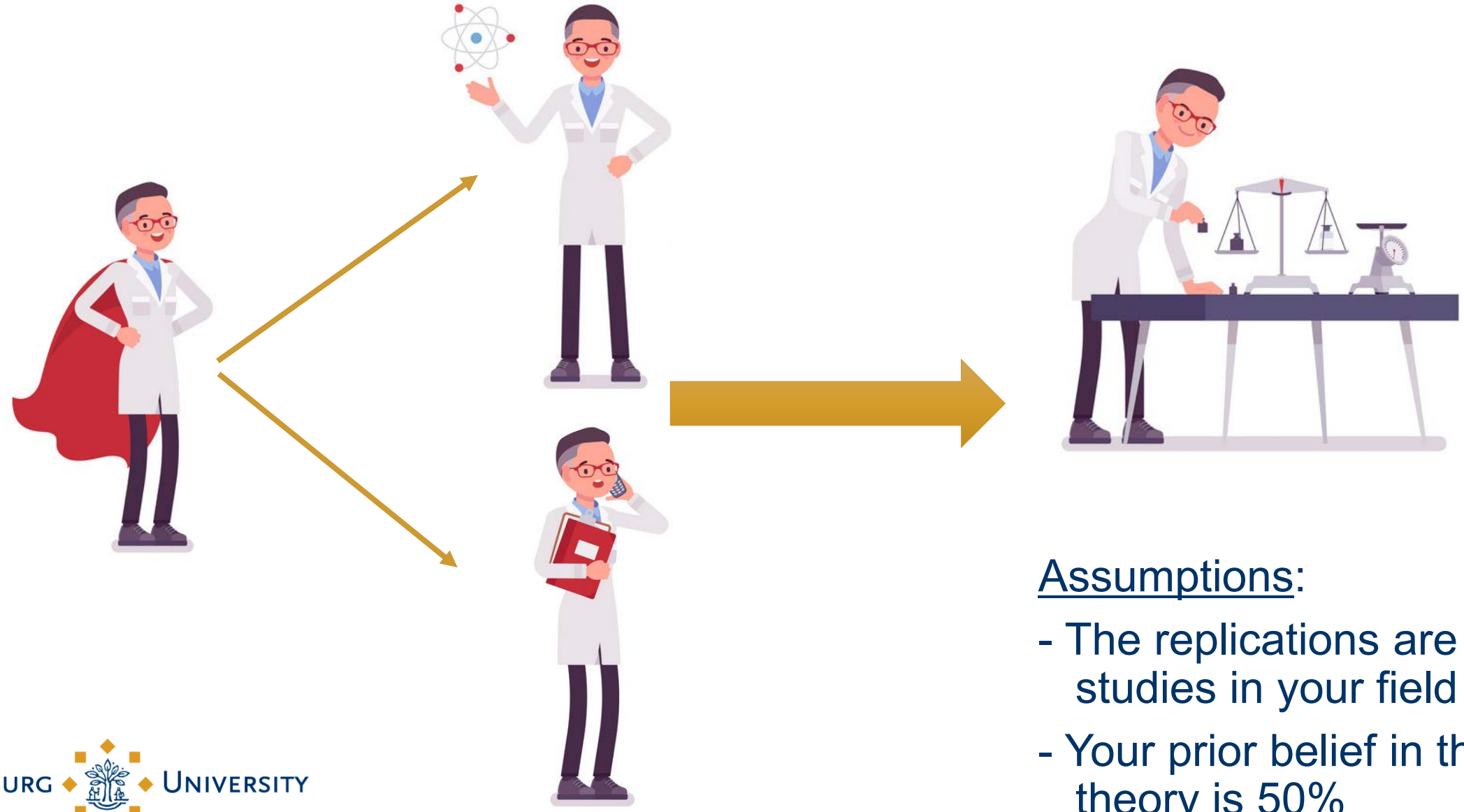
The Experiment

The Experiment

The Experiment

The Experiment

The Experiment



<u>Assumptions</u>:

- The replications are typical studies in your field
- Your prior belief in the theory is 50%

What are your beliefs?

Additional assumption:

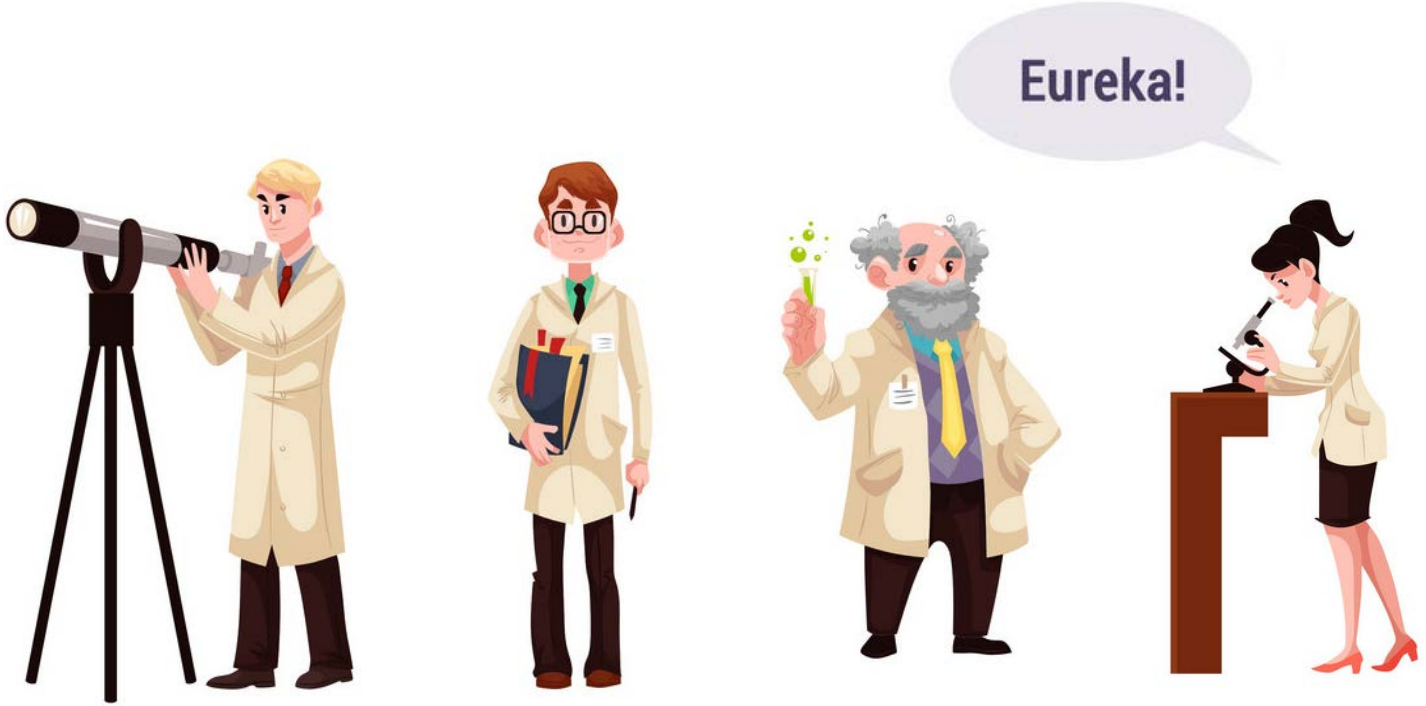
- The power of all four studies is 0.50

1. 19%
2. 39%
3. 59%

What are your beliefs?

Additional assumption:

- The power of all four studies is 0.50



1. 19%
2. 39%
3. 59%

What are your beliefs?

Additional assumption:

- The power of all four studies is 0.50

1. 77%
2. 87%
3. 97%

What are your beliefs?

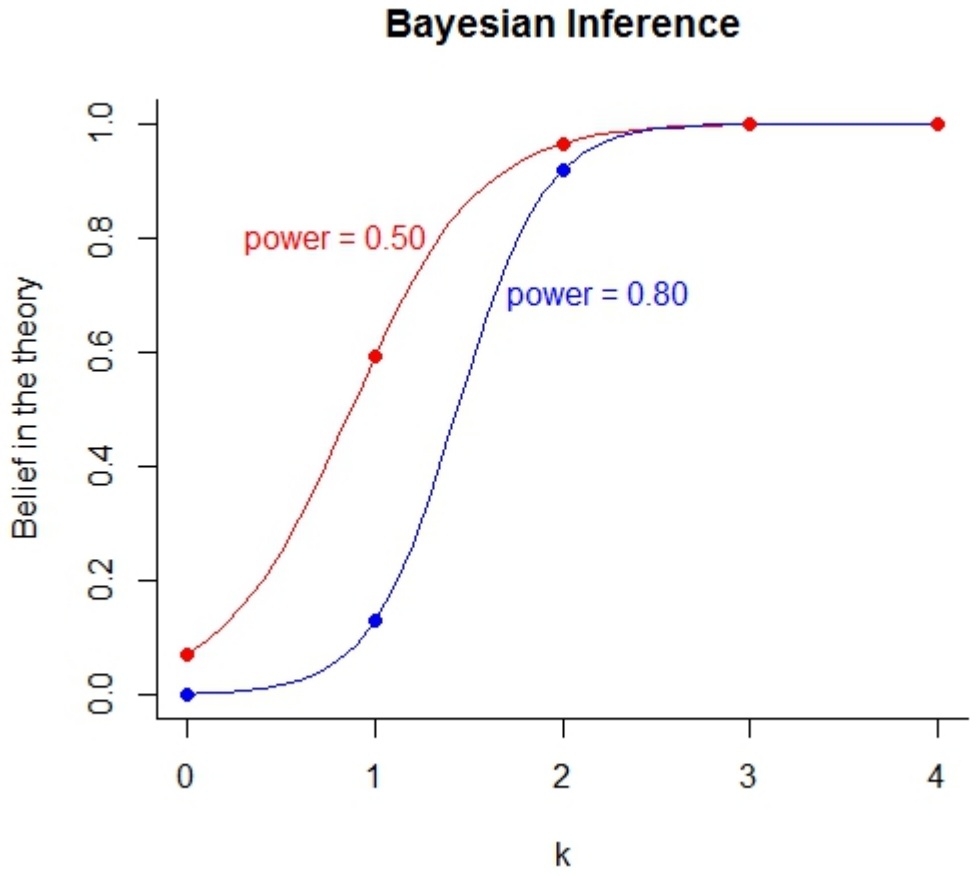
1. 77%
2. 87%
3. 97%

<u>Additional assumption</u>:

- The power of all four studies is 0.50

Bayesian Inference

$$p(H_A|data) = \frac{(1-\beta)^k \beta^{(n-k)}}{(1-\beta)^{k}\beta^{(n-k)} + \alpha^{n}(1-\alpha)^{n-k}}$$



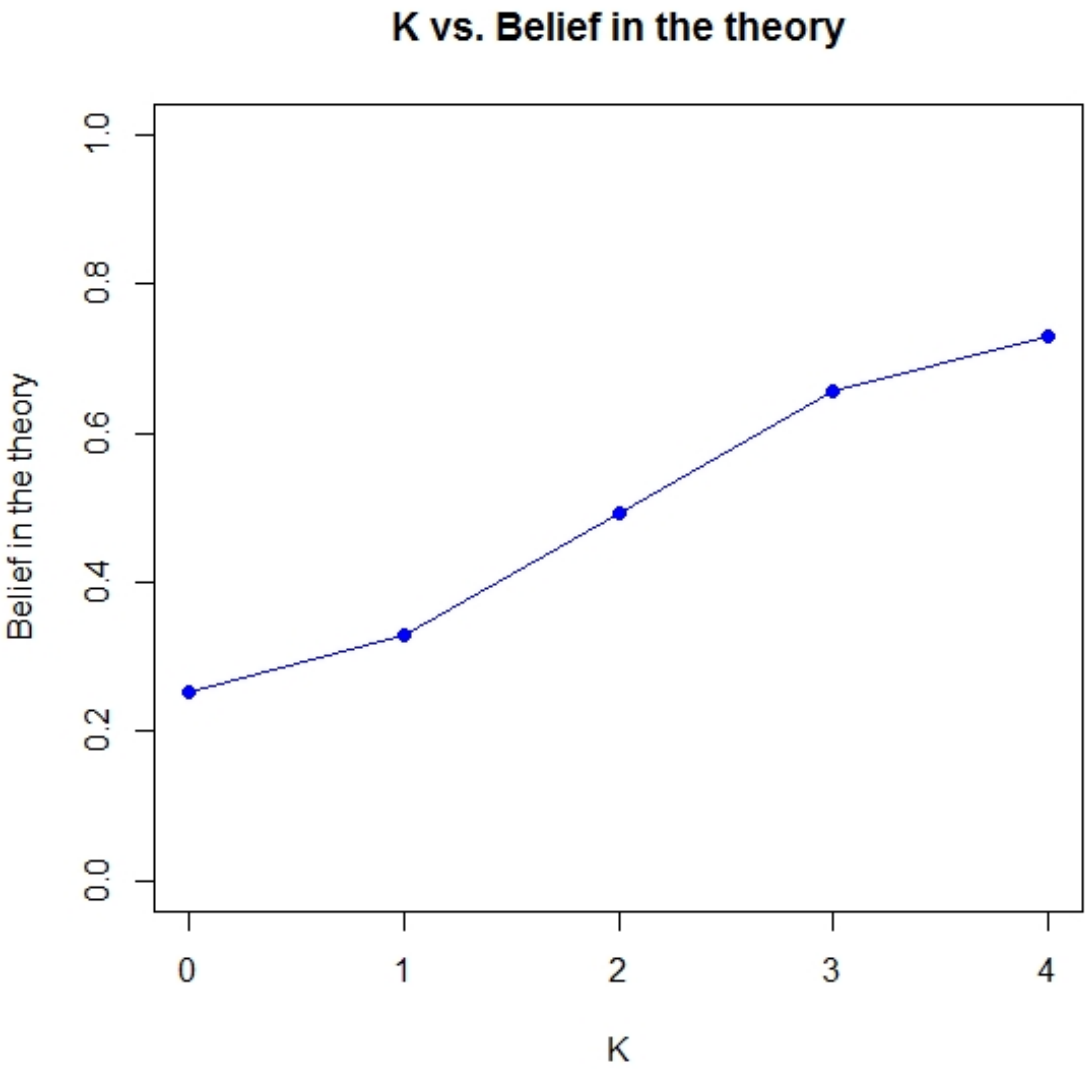
Main Hypotheses

1. There is a positive association between K and belief in the theory

2. Conceptual replications are valued more than direct replications.

3. ‘Researchers’ differ from ‘reviewers’ in their propensity to submit (for researchers) or recommend to submit (for reviewers) the set of studies for publication

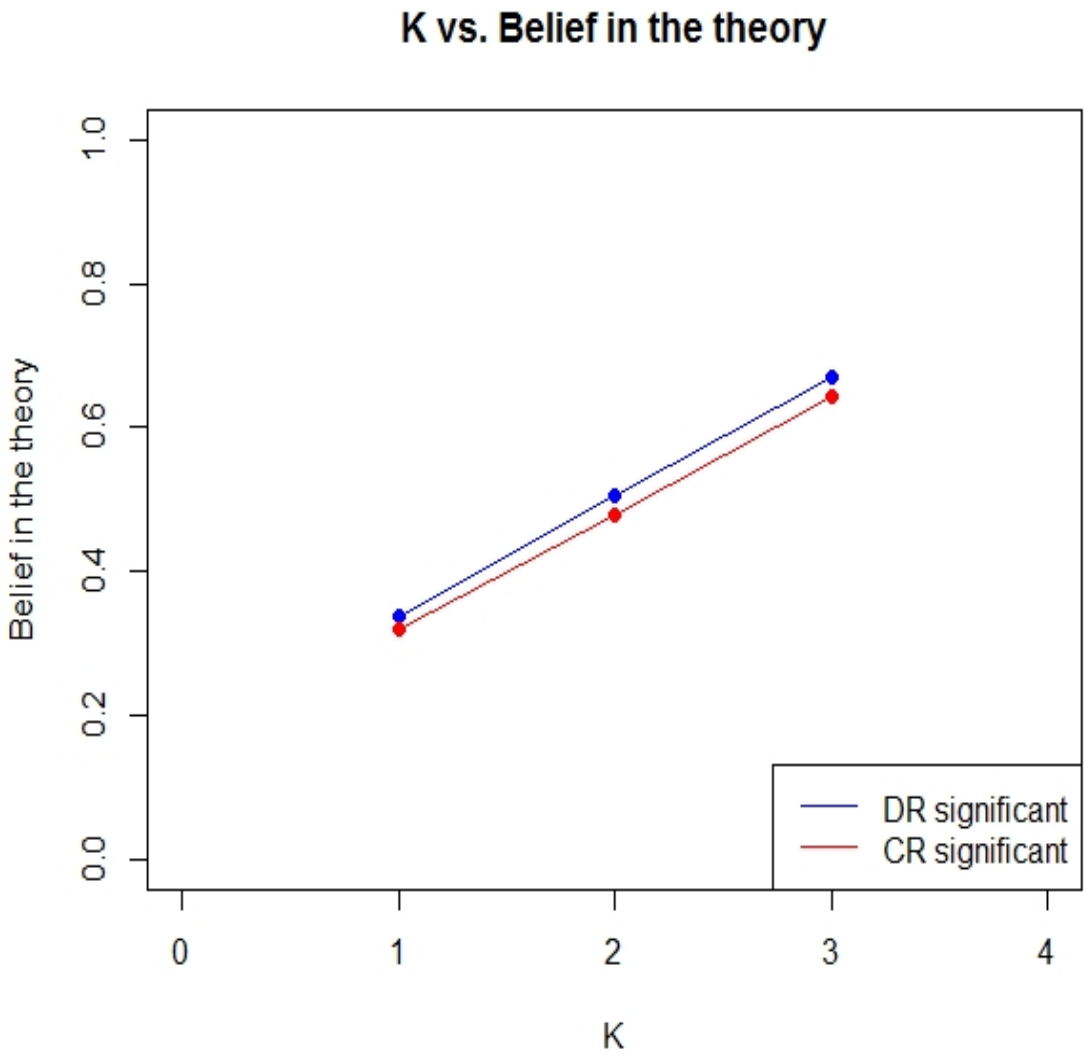
Main Results – Hypothesis 1



Finding 1:

There is a positive association between K and belief in the theory

Main Results – Hypothesis 2

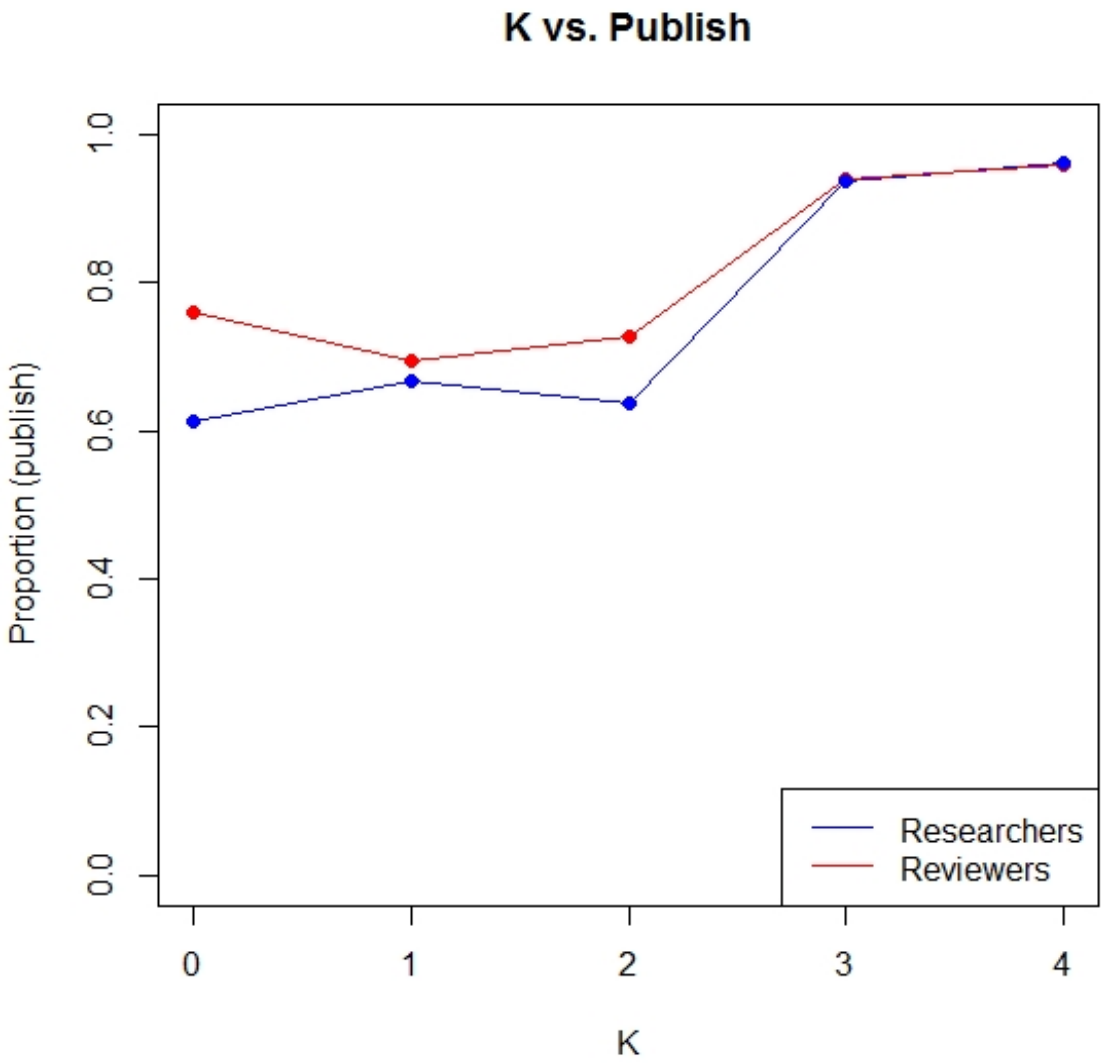


Finding 2:

Direct replications are valued (slightly) more than conceptual replications

Very small effect!

Main results – Hypothesis 3



Finding 3:

There is no difference between researchers and reviewers in their desire to publish

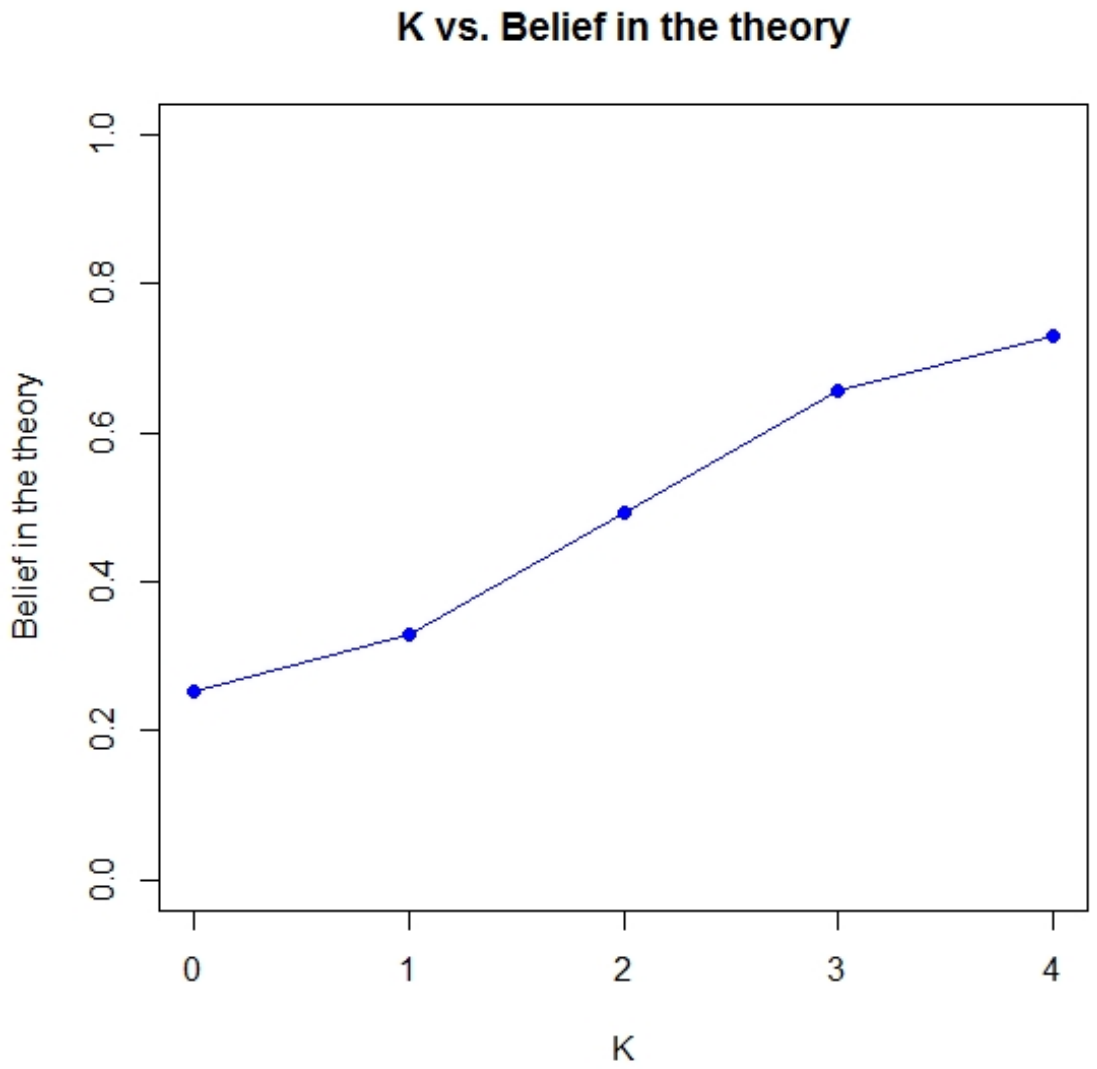
Main Results

1. Positive association between K and belief in the theory

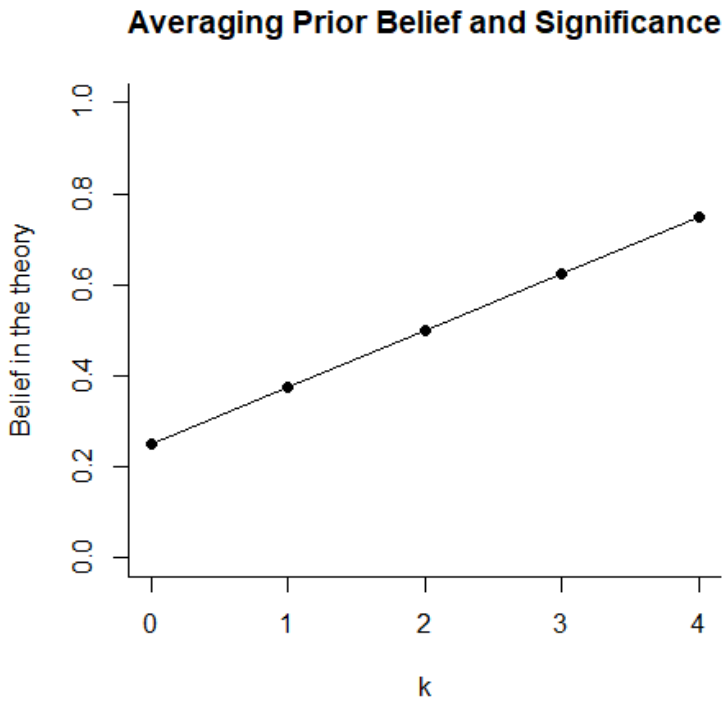
2. Direct replications are valued as much as conceptual replications

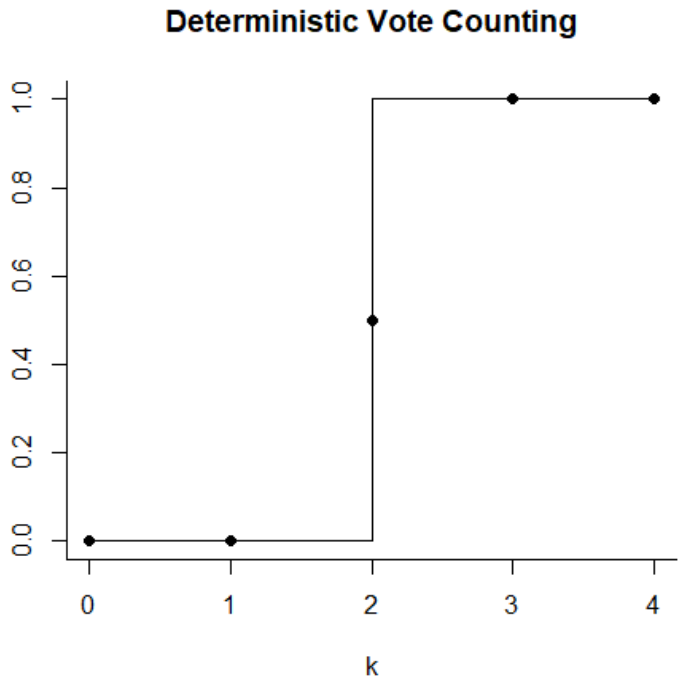
3. ‘Researchers’ do not differ from ‘reviewers’ in their propensity to submit or recommend the set of studies for publication

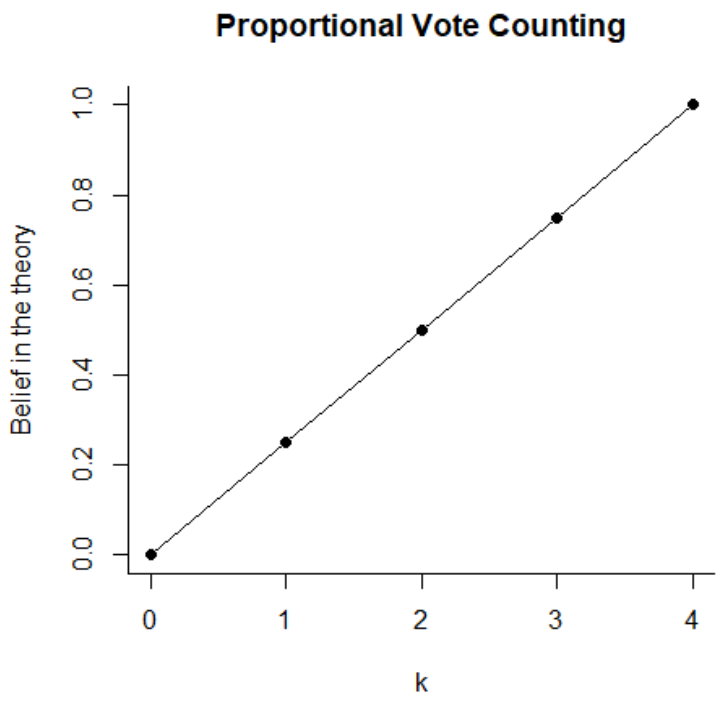
Main Results



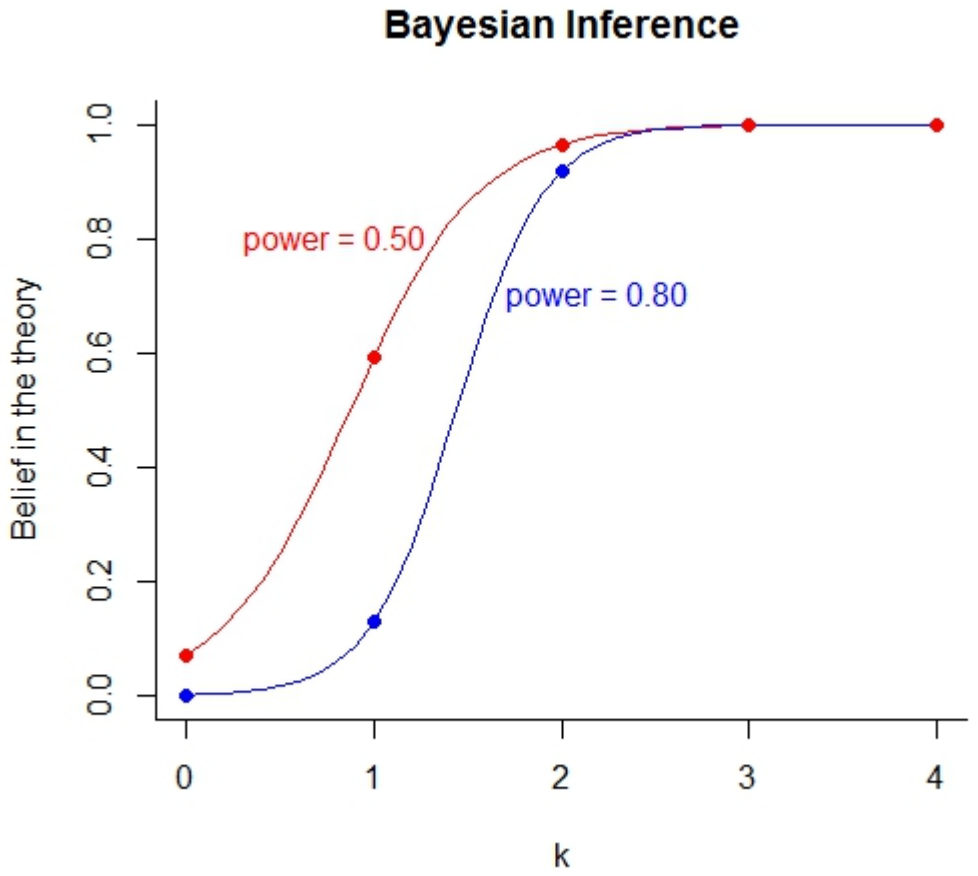
Other Heuristics







The normative heuristic



A Bayesian Approach

$$P(H_1|X) = \frac{\frac{1}{4} * P(X|H_1)}{\frac{1}{4} * P(X|H_1) + \frac{1}{4} * P(X|H_2) + \frac{1}{4} * P(X|H_3) + \frac{1}{4} * P(X|H_4)}$$

where $P(X|H_i)$ is the likelihood of the data given heuristic *i*

Categorization

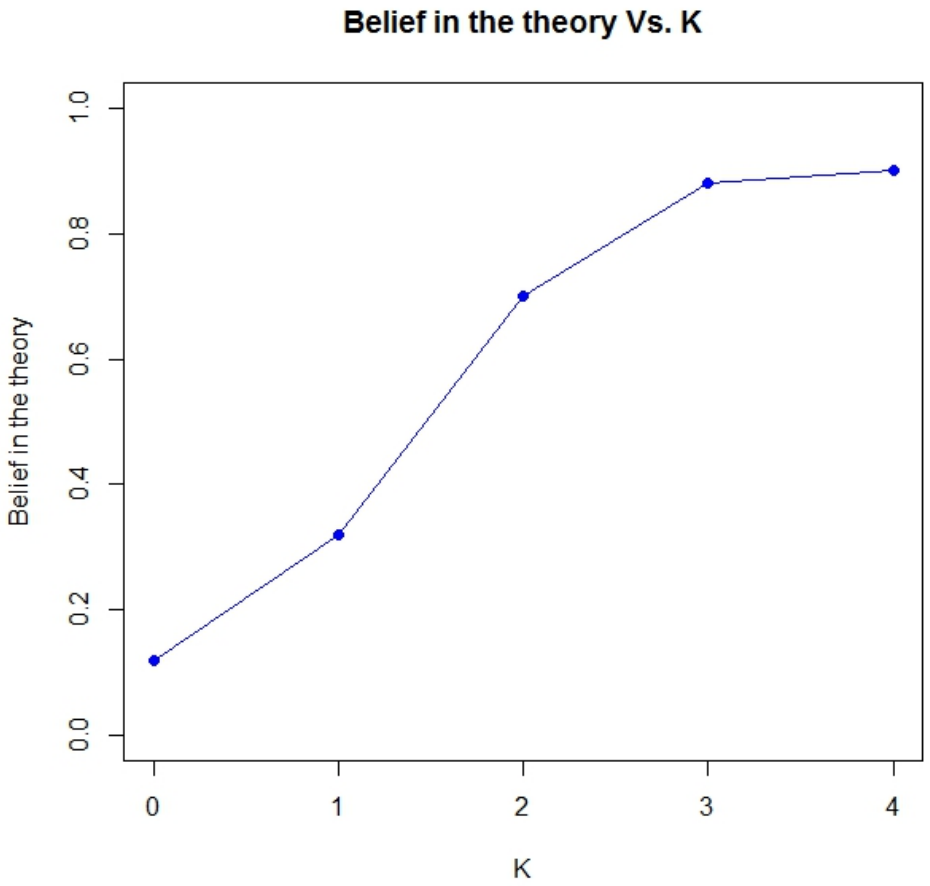
Strict:

1) The heuristic should be more than three times as likely as the other heuristics combined (Bayes Factor > 3)
2) The heuristic should outperform a benchmark heuristic ($RMSE_{\ Heuristic} < RMSE_{\ Benchmark}$)

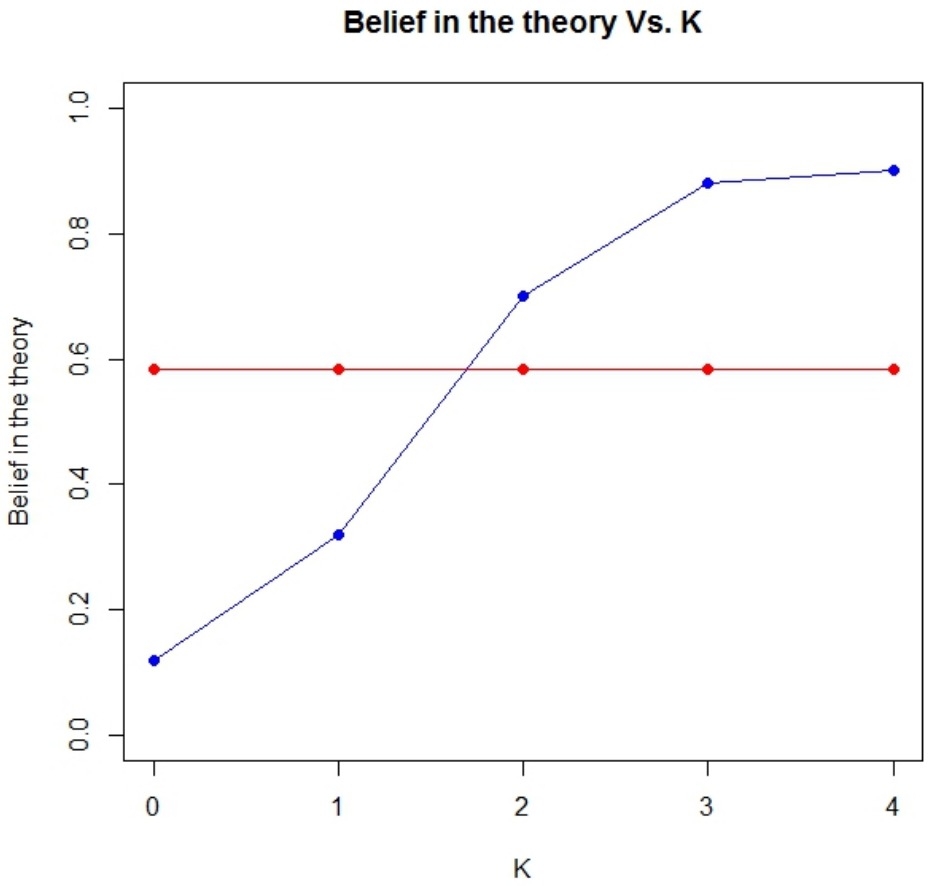
Lenient:

1) The heuristic should be more likely than the other heuristics
2) The heuristic should outperform a benchmark heuristic ($RMSE_{\ Heuristic} < RMSE_{\ Benchmark}$)

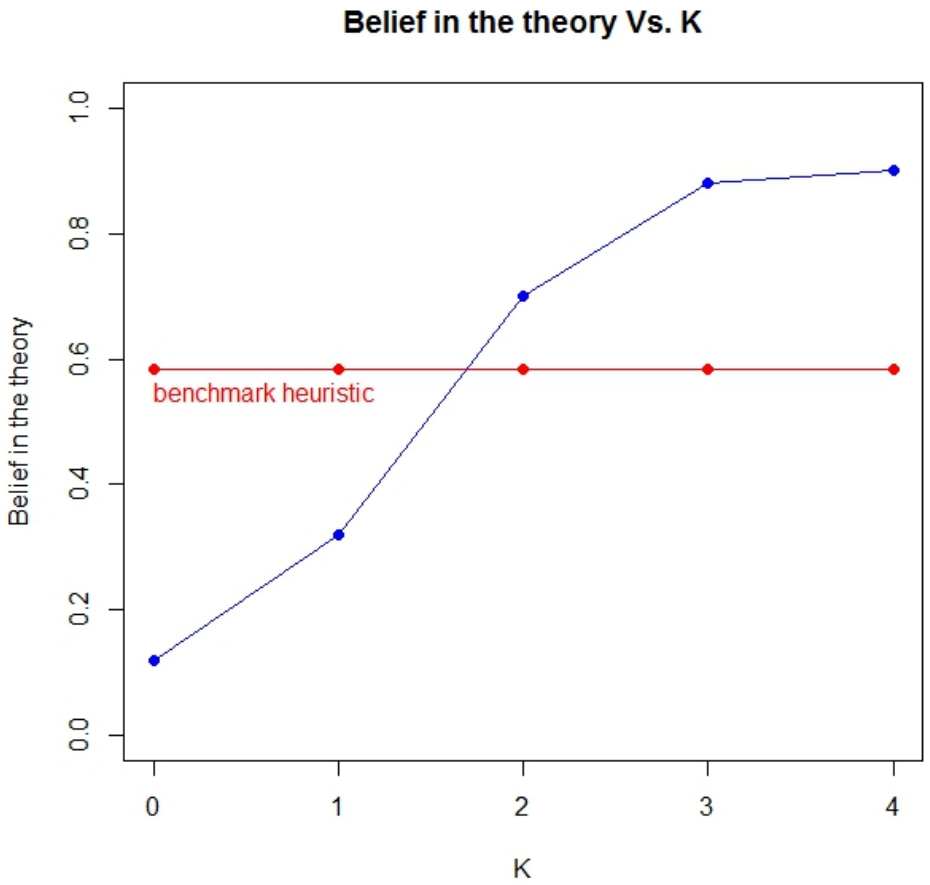
The Benchmark Heuristic



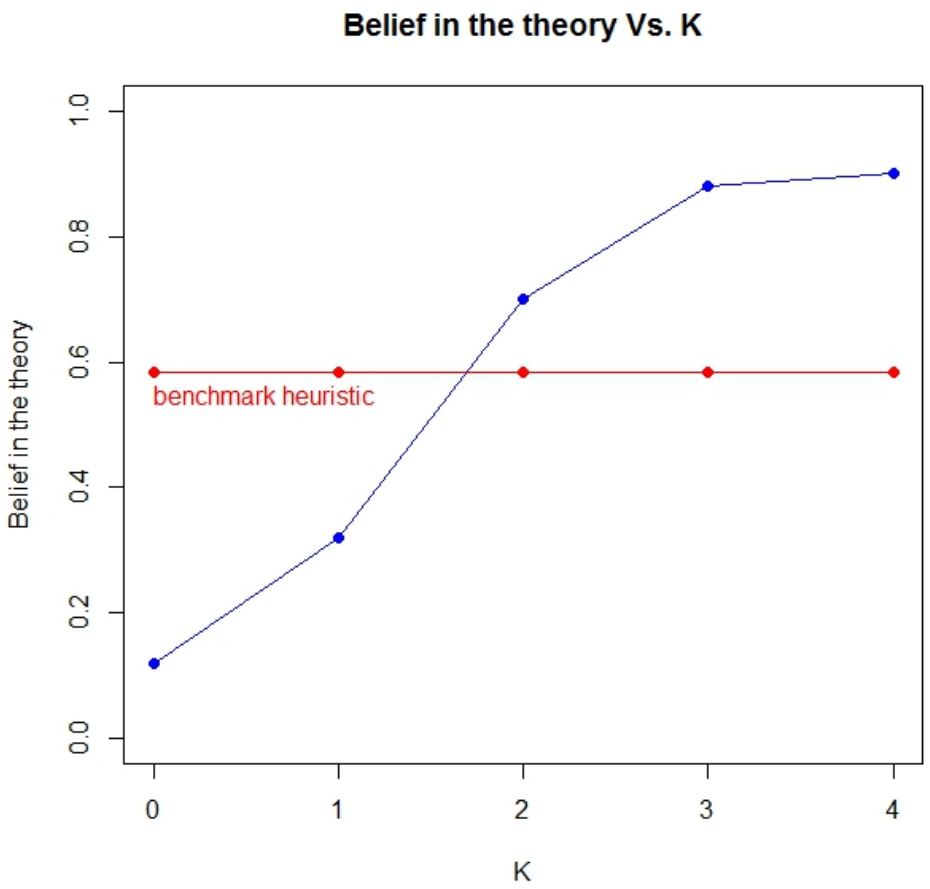
The Benchmark Heuristic

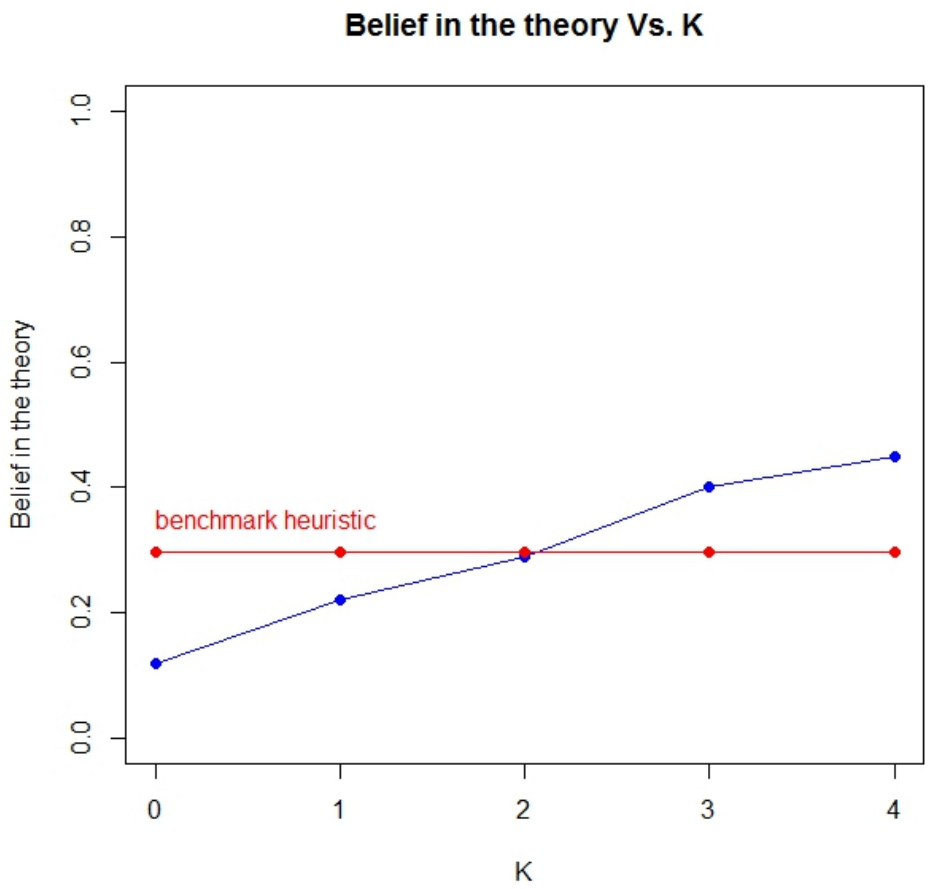


The Benchmark Heuristic

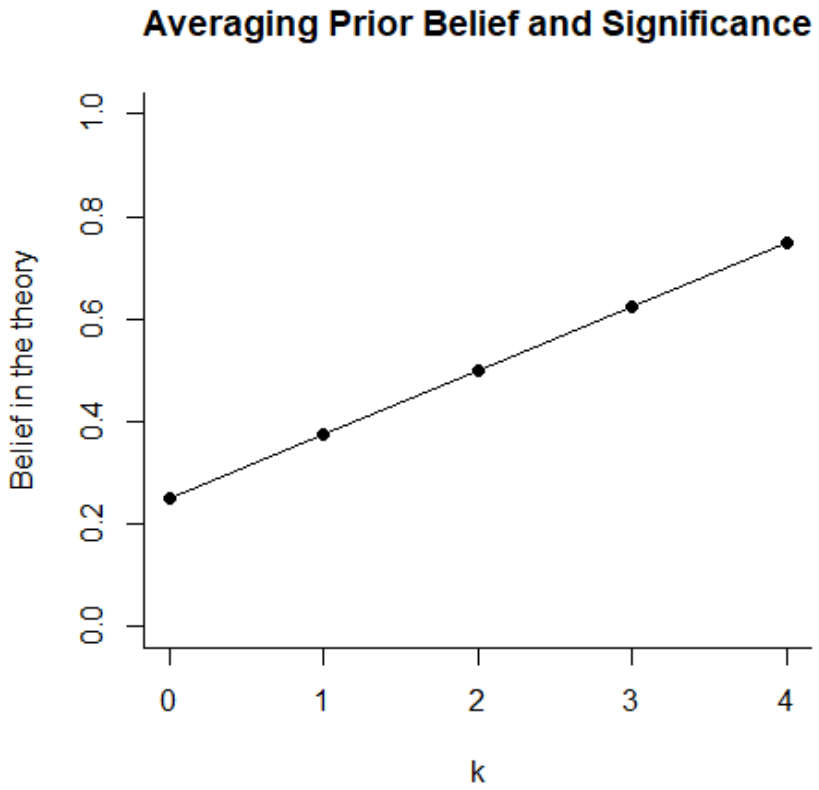


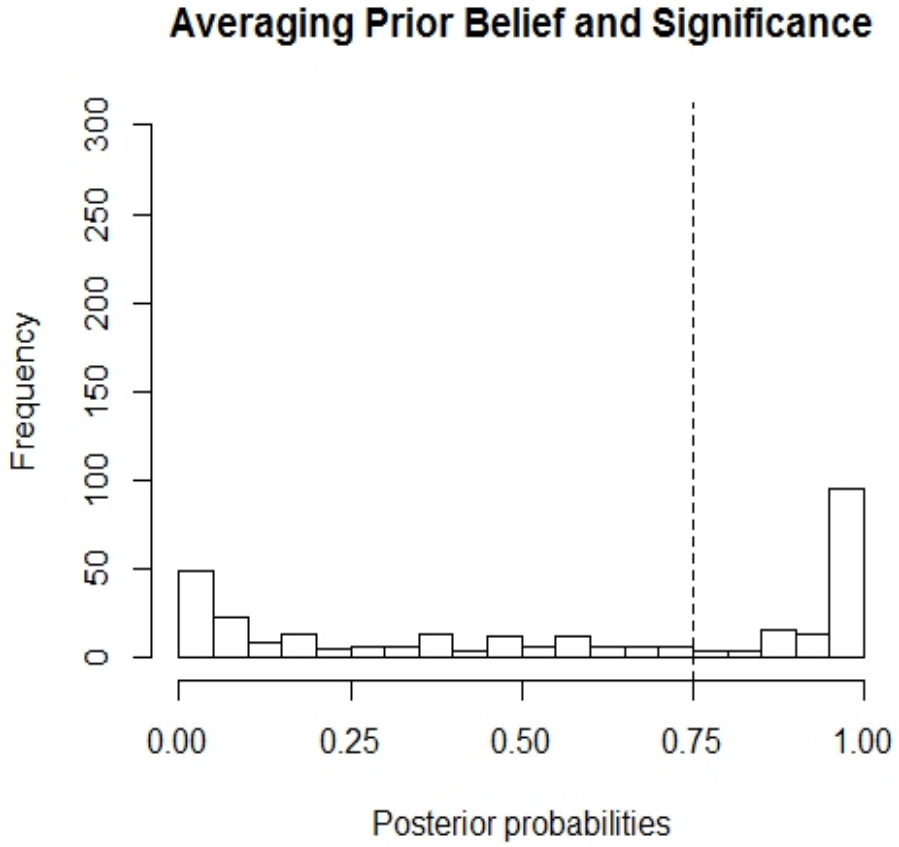
The Benchmark Heuristic



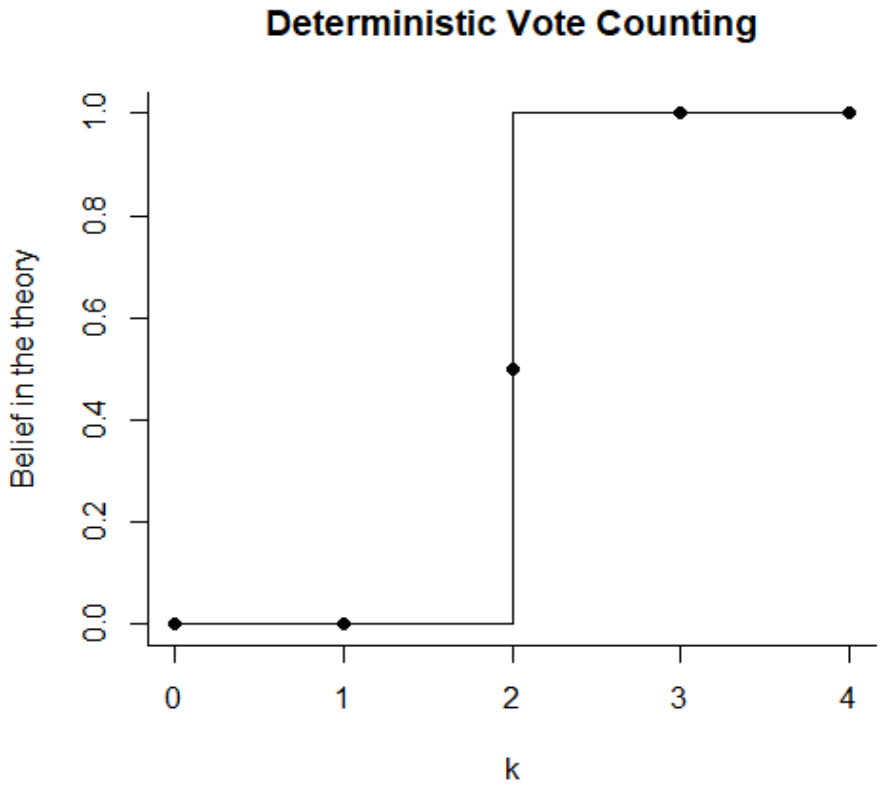


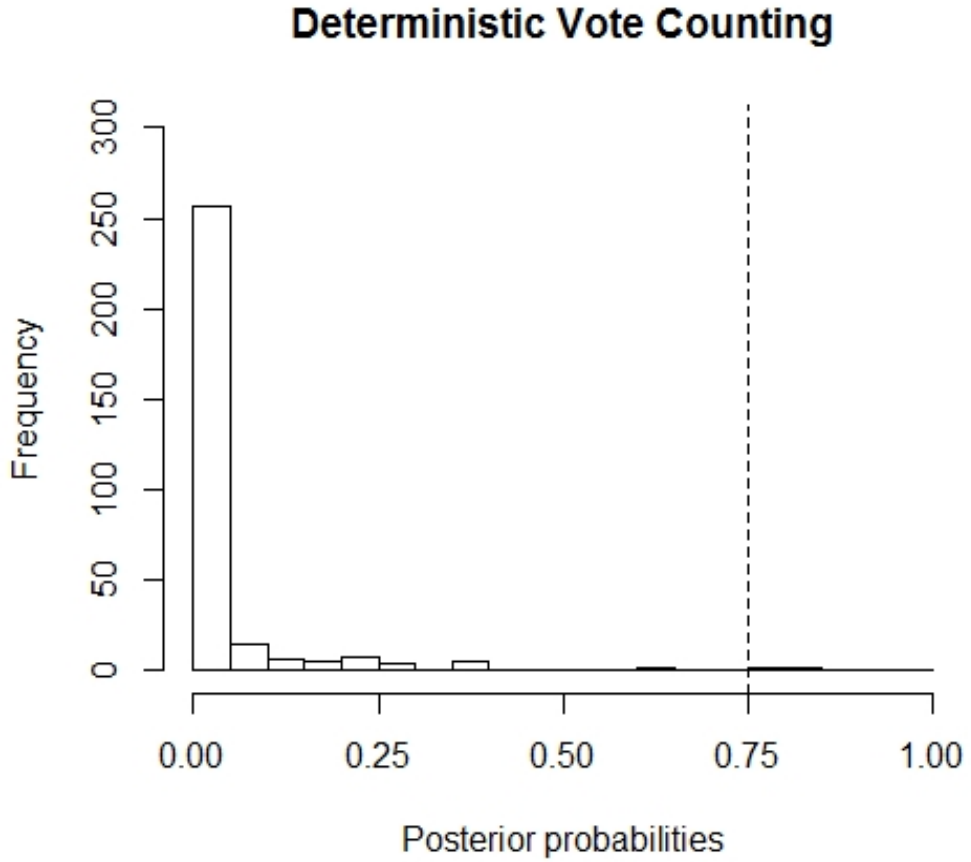
The Posterior Probabilities



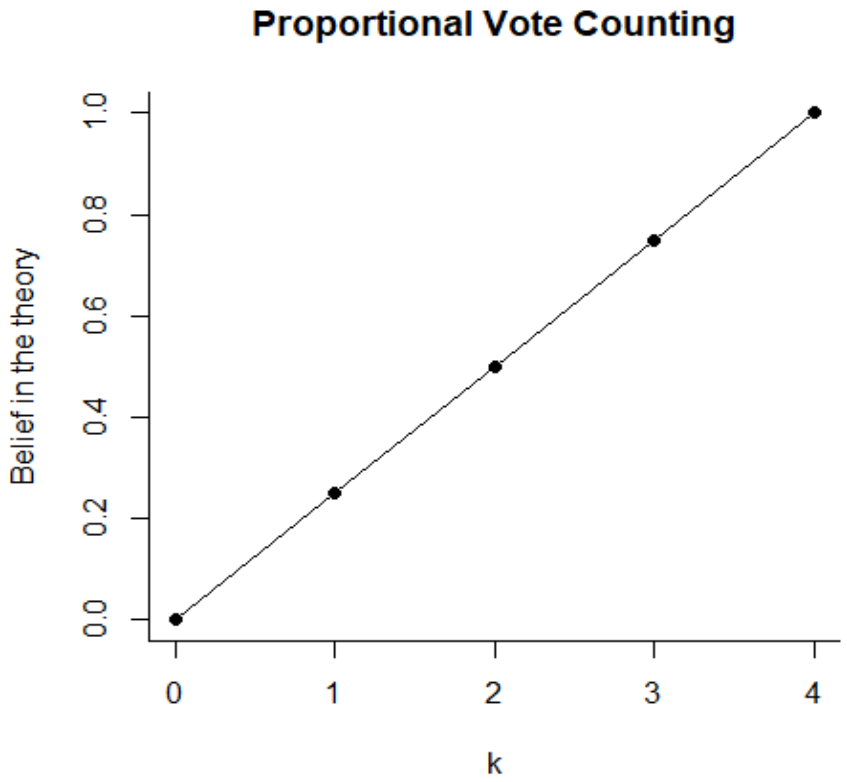


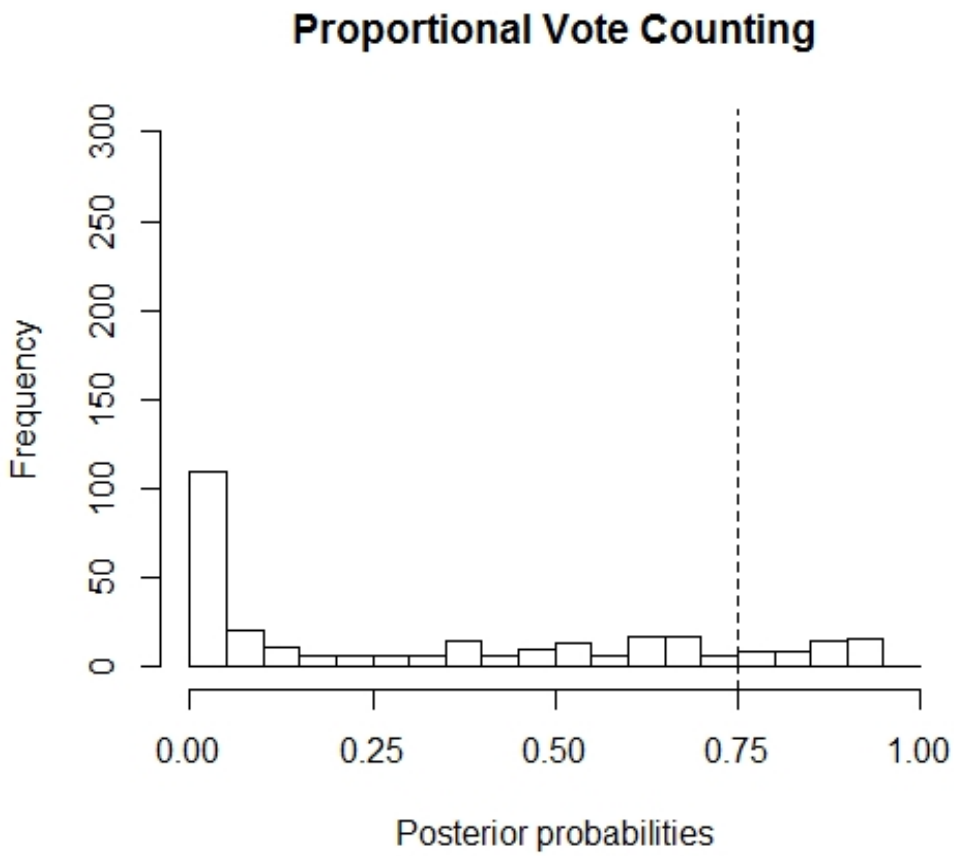
The Posterior Probabilities



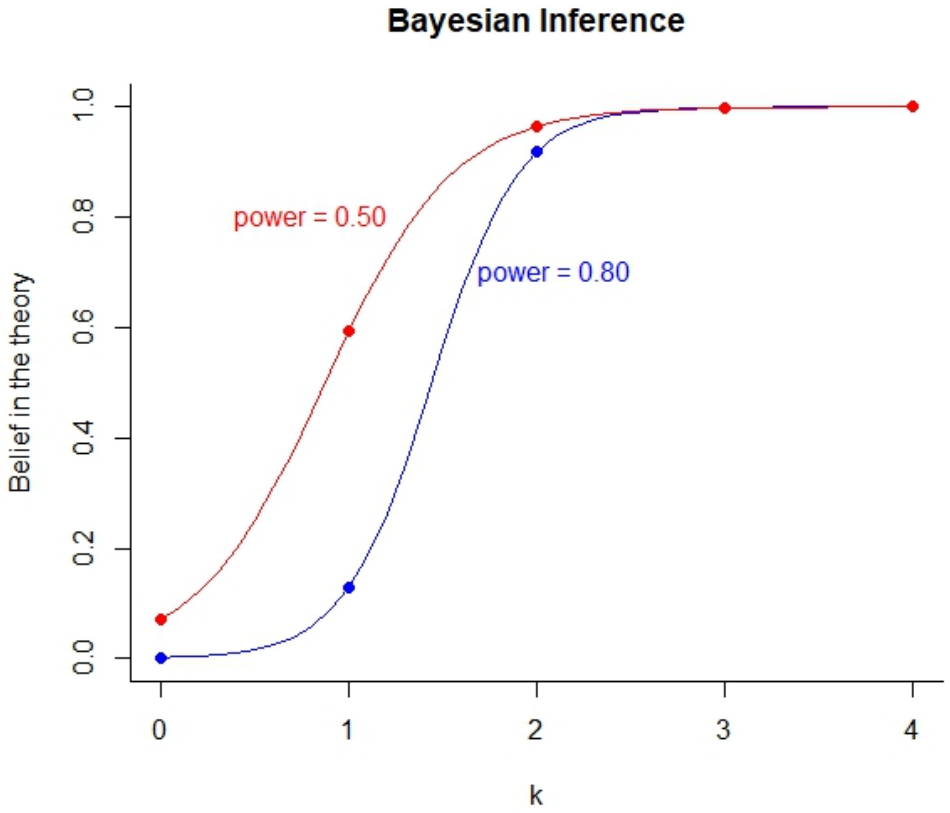


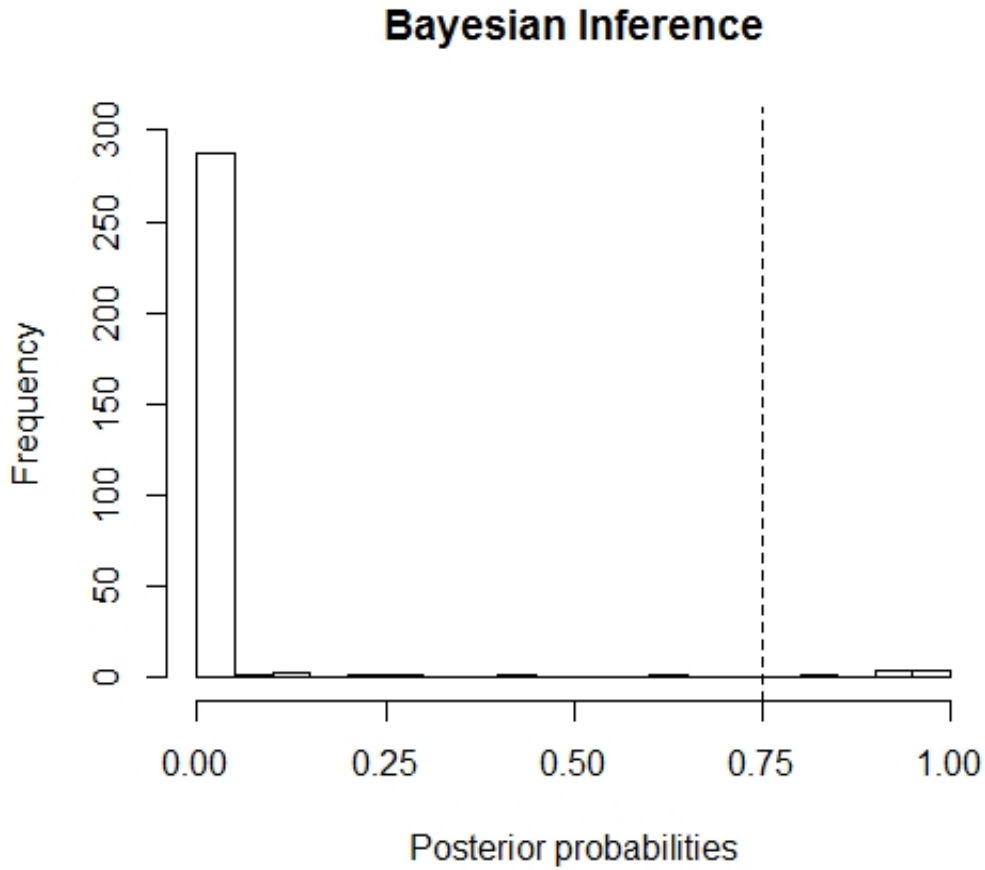
The Posterior Probabilities



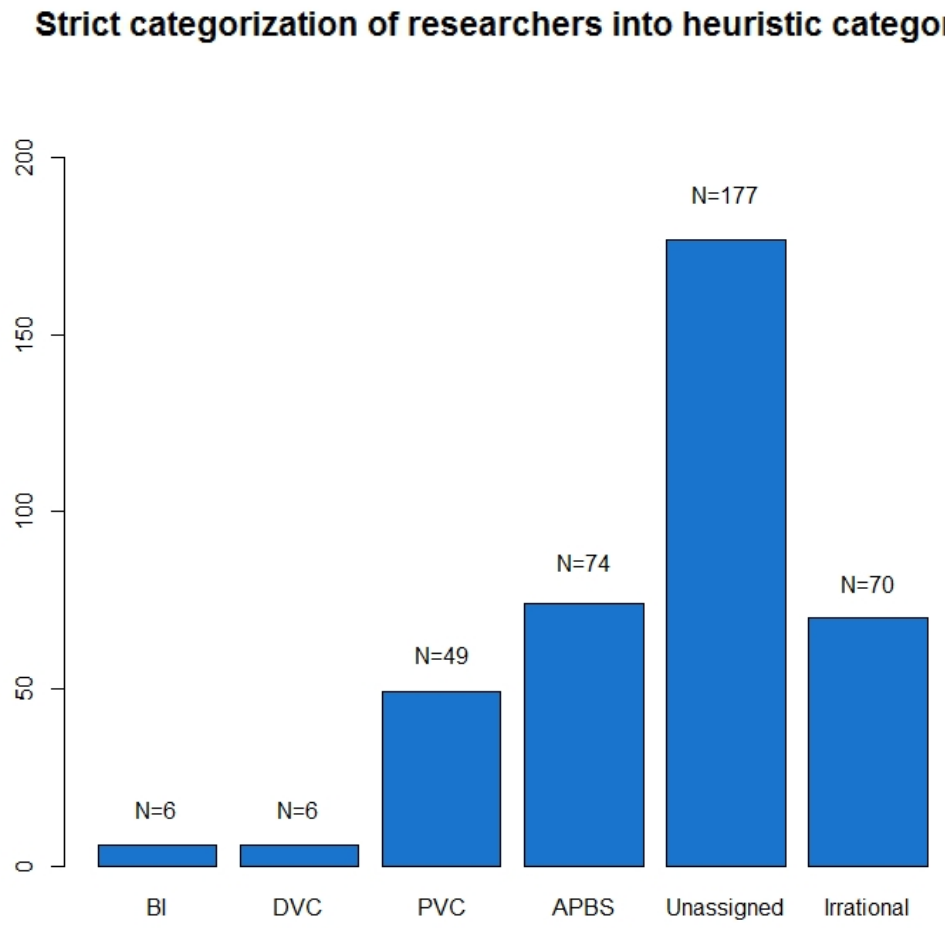


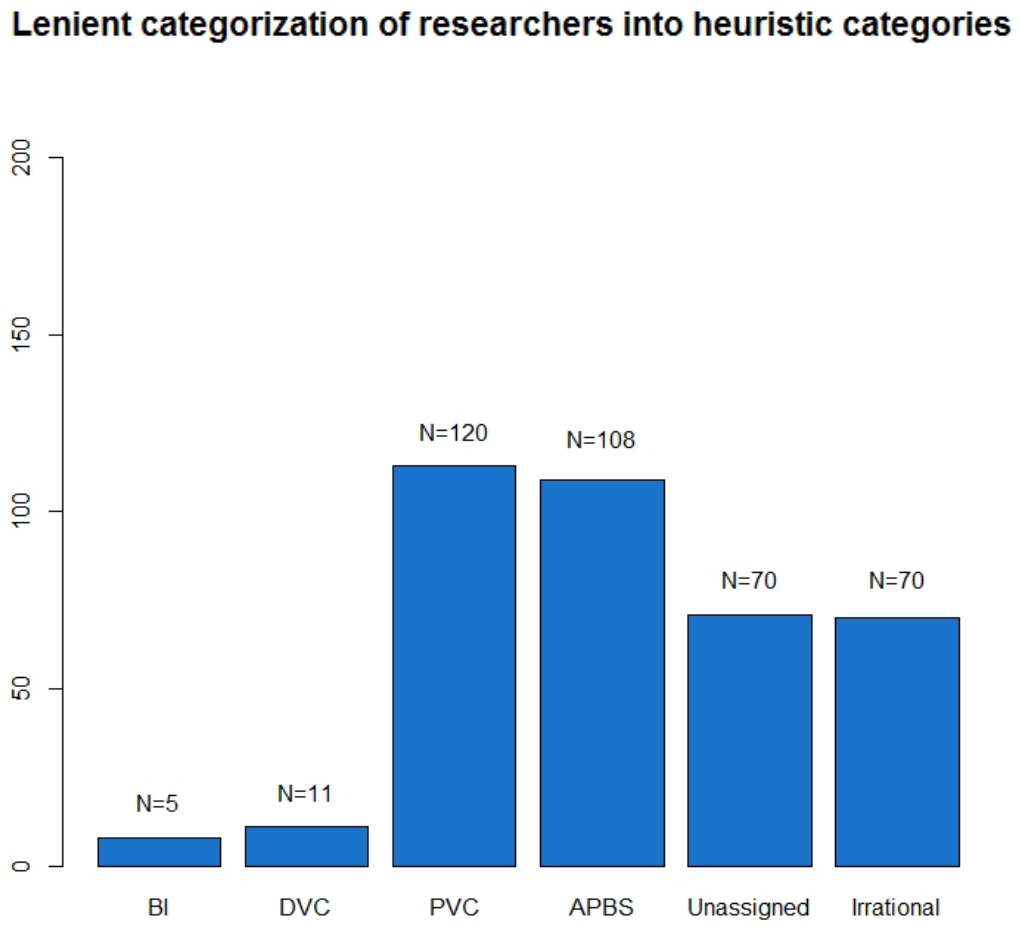
The Posterior Probabilities





The Results

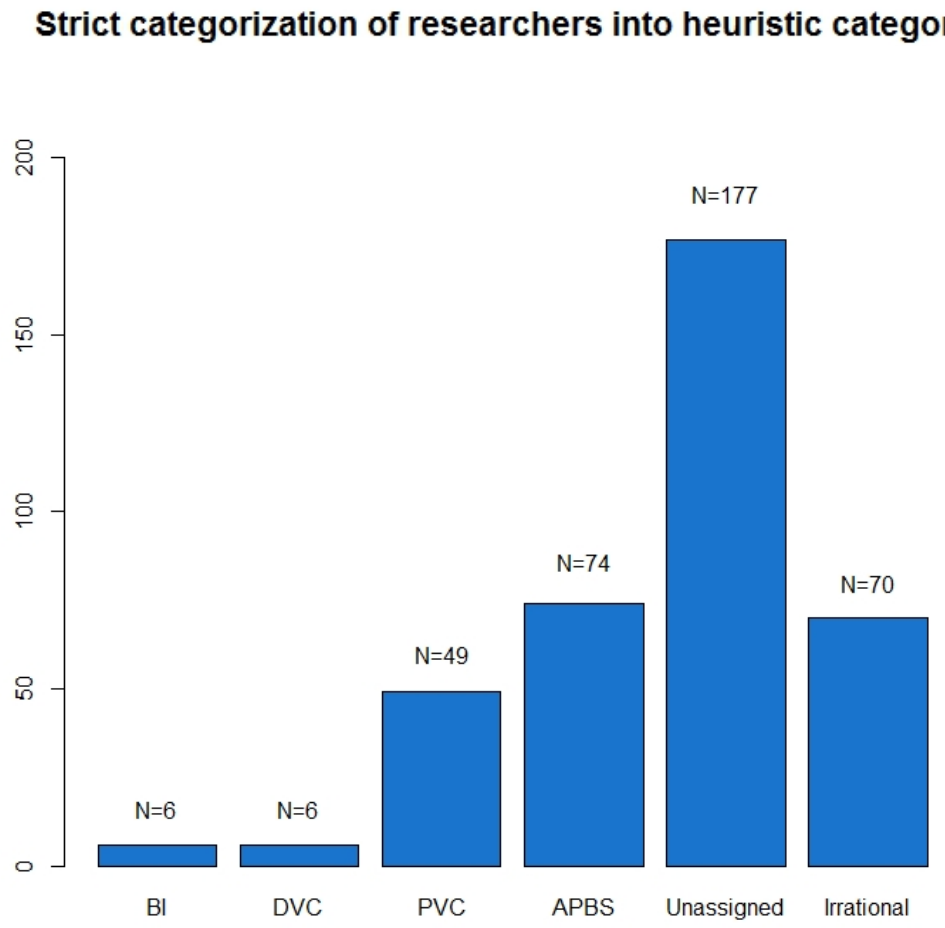


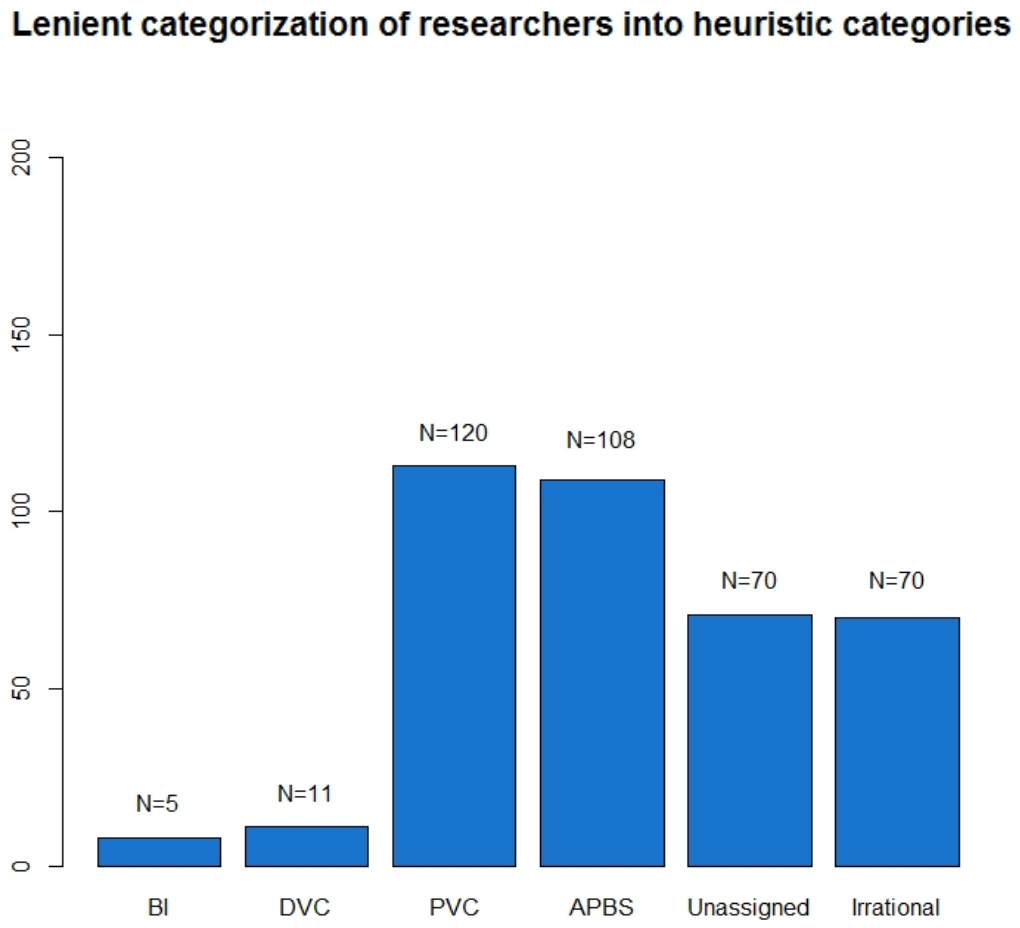


Finding 4:

Almost no researchers use ‘Bayesian inference’ and ‘deterministic vote counting’

The Results

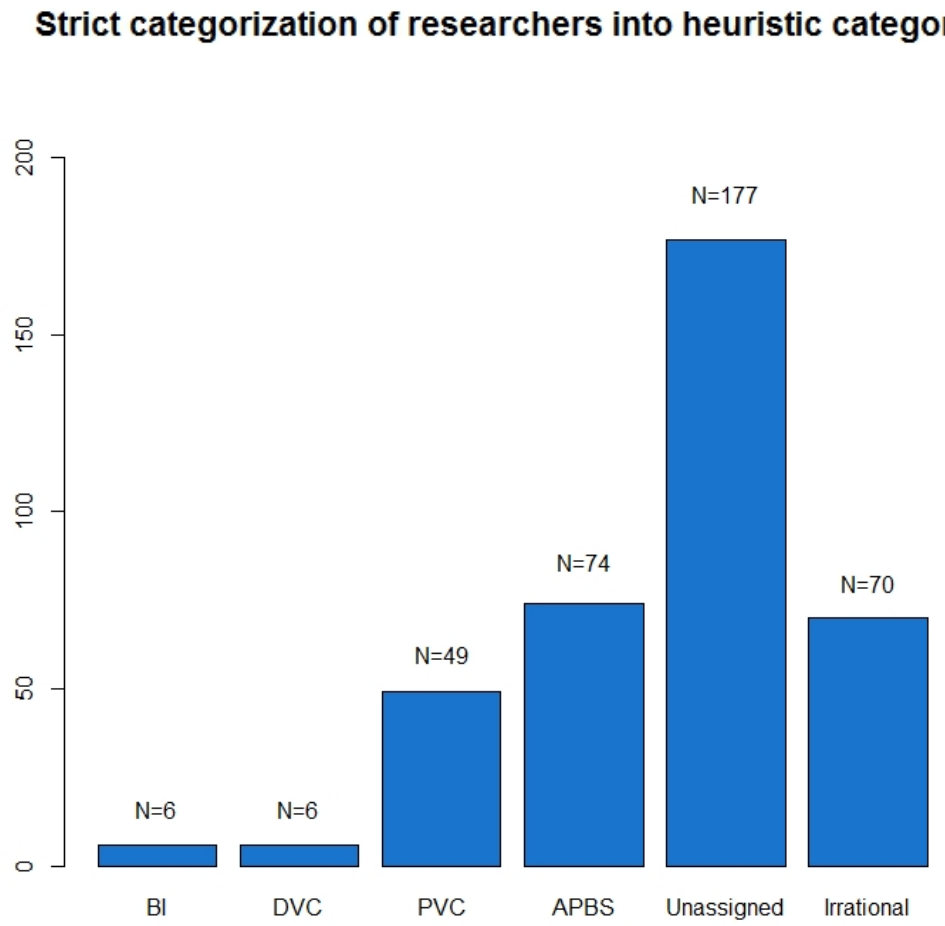


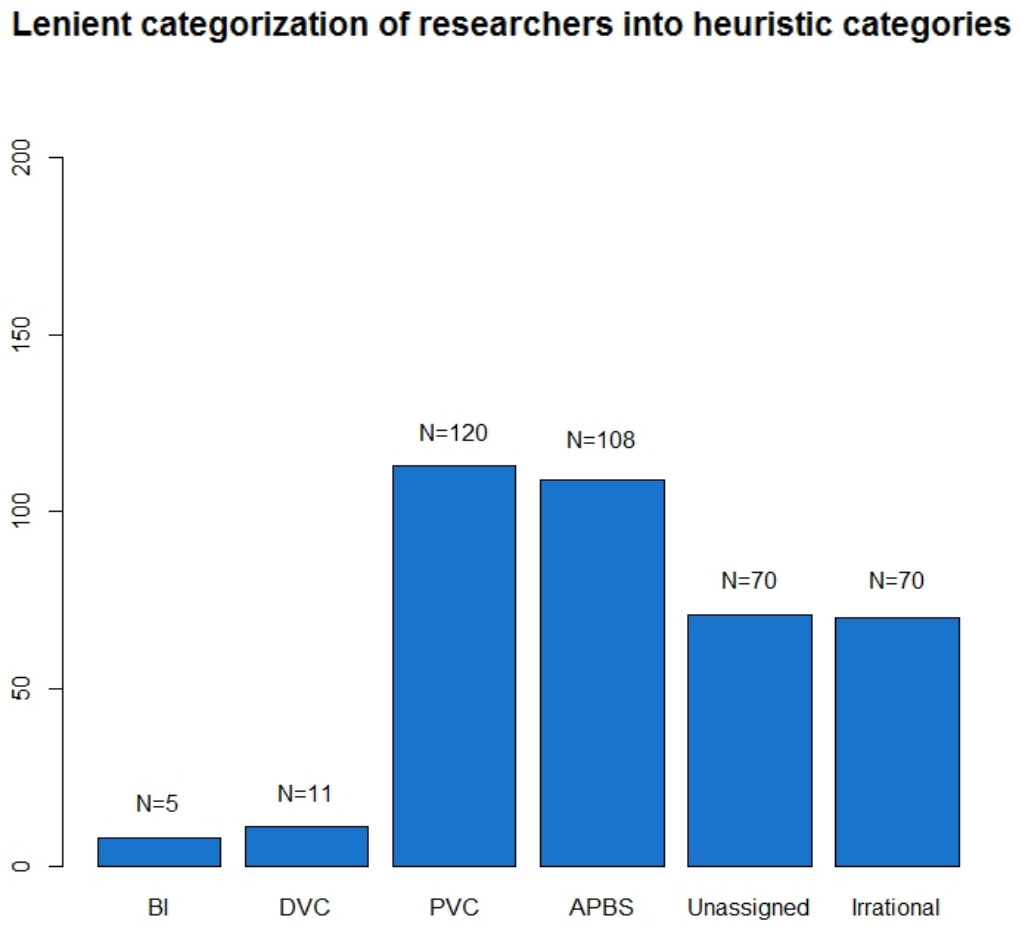


Finding 5:

A significant amount of researchers use ‘proportional vote counting’ and ‘averaging prior belief and significance’

The Results





Finding 6:

A significant amount of researchers are classified as irrational

Conclusions

- Researchers often use simple decision rules to make sense of statistical results
- Researchers lack statistical intuition
- We need to educate researchers/educators
- Meta-education?

Questions

Contact:

o.r.vdnakker@uvt.nl

@DenOlmo

Preprint:

https://psyarxiv.com/xyks4

Data and materials:

https://osf.io/2g4wf