

Open Science
in

Computational
Social
Science

Prof. Dr. Alexander Wuttke

DATA IN POLITICS



What is science?

Why trust science?



Group work

Take 10 minutes to discuss in groups

- **What is science?**

(How can you identify a knowledge claim based on science?
How can you discern science from non-science?)

- **Why trust science?**

(What makes scientific knowledge claims more trustworthy than other knowledge claims (e.g. based on tradition, religion, authority, experience, journalism, industry insights)



It is clear then, that the **idea of a fixed method**, or a fixed theory of rationality, rests on **too naïve a view of man and his social surroundings**. To those who look at the rich material provided by history, and who are not intent on impoverishing it in order to please their lower instincts, their craving for intellectual security in the form of clarity, precision, objectivity,” [and] “truth,” it will become clear that there is only one principle that can be defended under all circumstances and in all stages of human development. It is the principle: **anything goes**.

Feyerabend



Science's social nature

and its epistemic nature



Science's social nature

and its epistemic nature



**Social learning
and vetting**



**Expert
consensus**

Requires openness to work!



Scientific Ethos (Merton)

- **Commun(al)ism**

- all scientists should have **common ownership of scientific goods** (intellectual property), to promote collective collaboration; secrecy is the opposite of this norm.

- **universalism**

- **scientific validity is independent of the sociopolitical status**/personal attributes of its participants.

- **disinterestedness**

- scientific institutions act for the benefit of a common scientific enterprise, **rather than for the personal gain** of individuals within them.

- **organized skepticism**

- scientific claims should be exposed to **critical scrutiny** before being accepted: both in methodology and institutional codes of conduct.

3 Problems

- Systemic incentive problems
- Idiosyncratic modeling choices
- Reproducibility



Systemic incentive problems



Incentive system and collective goods

Things you
need to do
to advance
your career

Things you
need to do
to advance
science

The daily practice of an elite scientist

Read the following Blog post by Stanford professor Brian Wansink

https://drive.google.com/file/d/180_DyYxV40z66eAQtkAVwIbjoF5G81Zq/view#

- Who was right or wrong when and why?
- Which practices are problematic?



Amazing!



Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

Keywords: psi, parapsychology, ESP, precognition, retrocausation





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Psychological Science
Volume 22, Issue 11, November 2011, Pages 1359-1366
© Association for Psychological Science 2011, Article Reuse Guidelines
<https://doi.org/10.1177/0956797611417632>



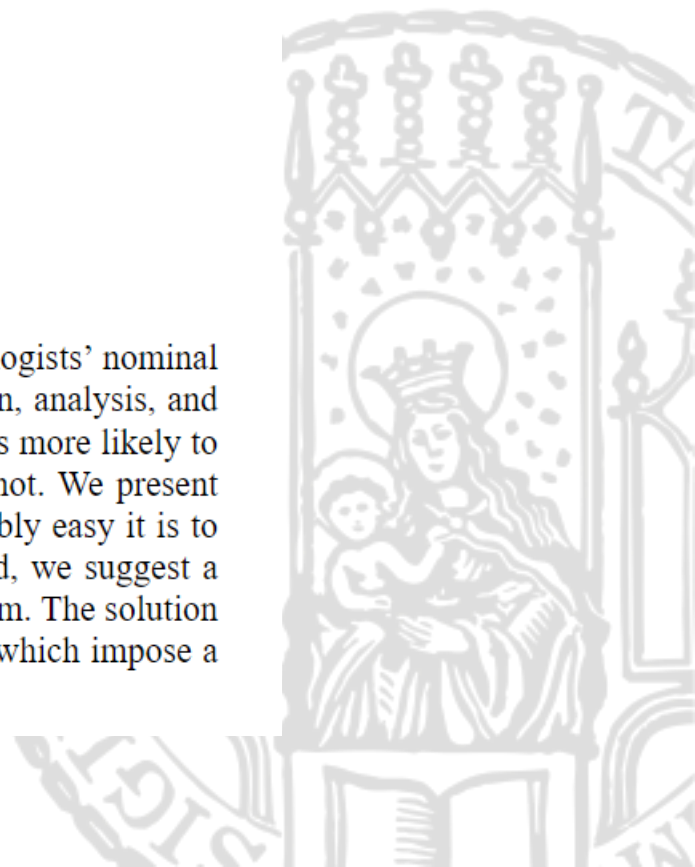
General Article

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

Abstract

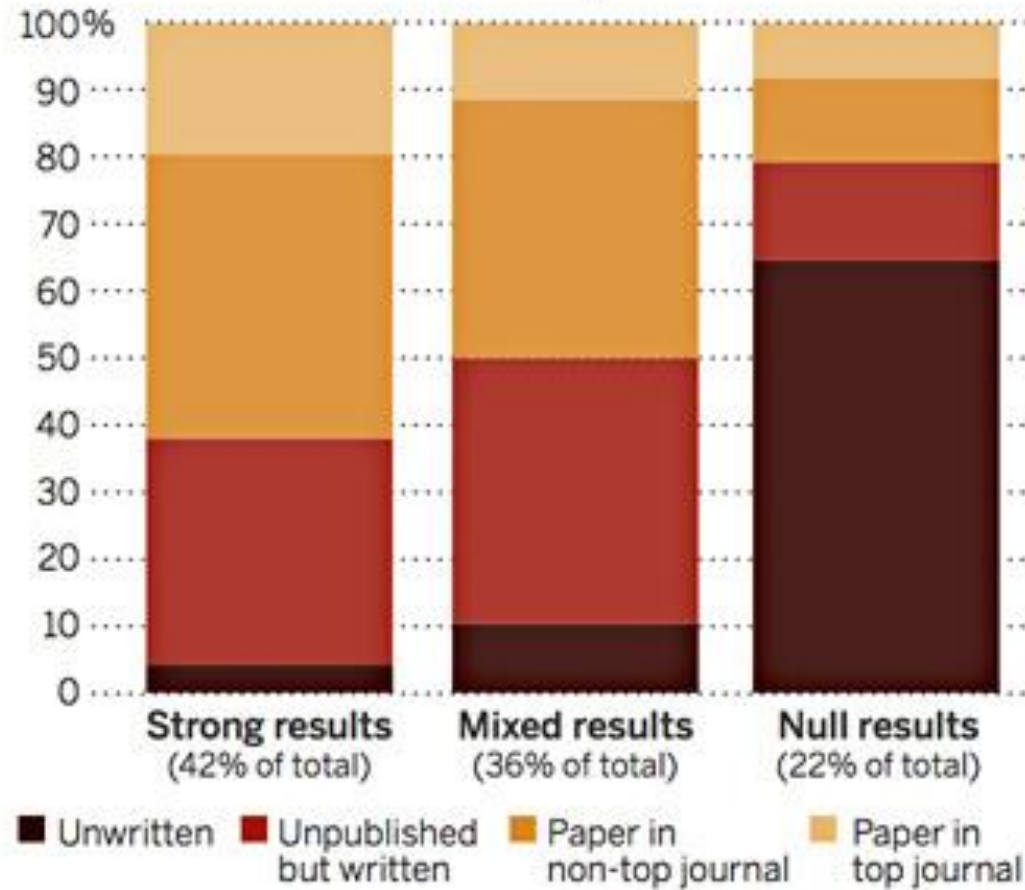
In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ($\leq .05$), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.



Publication Bias

Most null results are never written up

The fate of 221 social science experiments



Source: A. Franco et al., Science (28 August)

29. Juli 2023

Source: Mother Jones

Analytical discretion

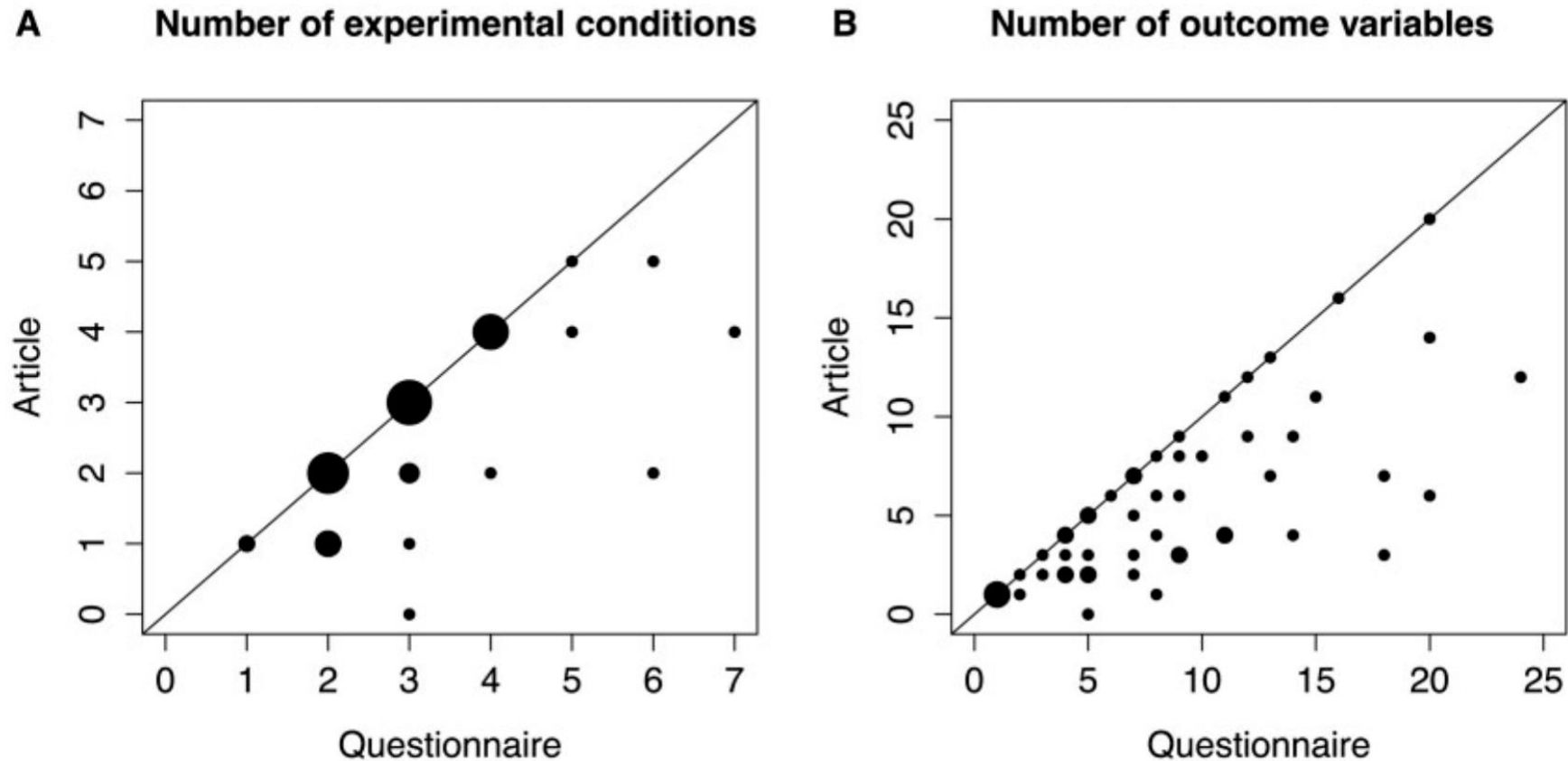
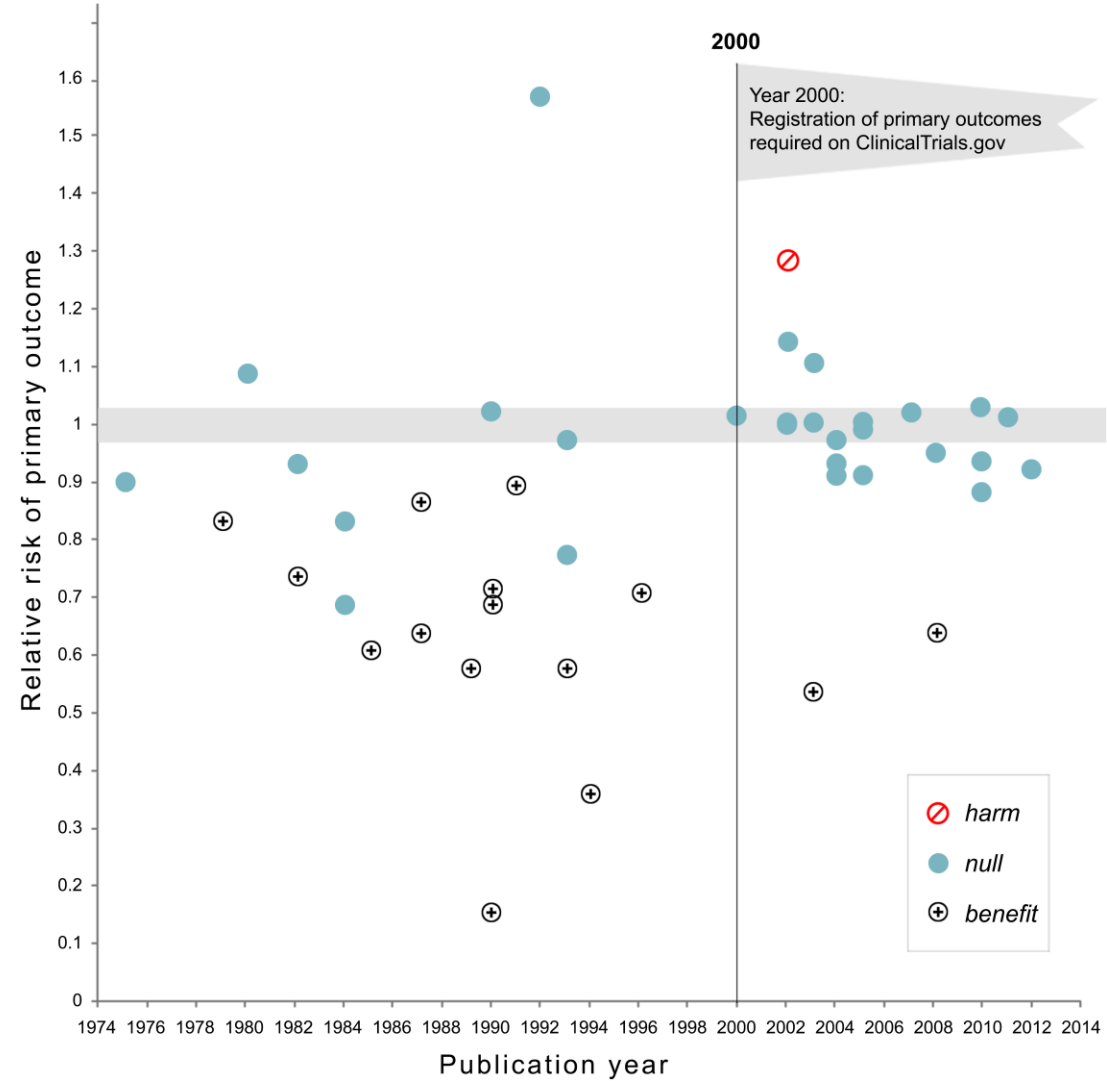


Fig. 1 Comparing design features in questionnaires and published results in TESS studies.
Note. Point size in each panel is proportional to the number of studies with a particular questionnaire-article value pair.

Franco et al. (2015)

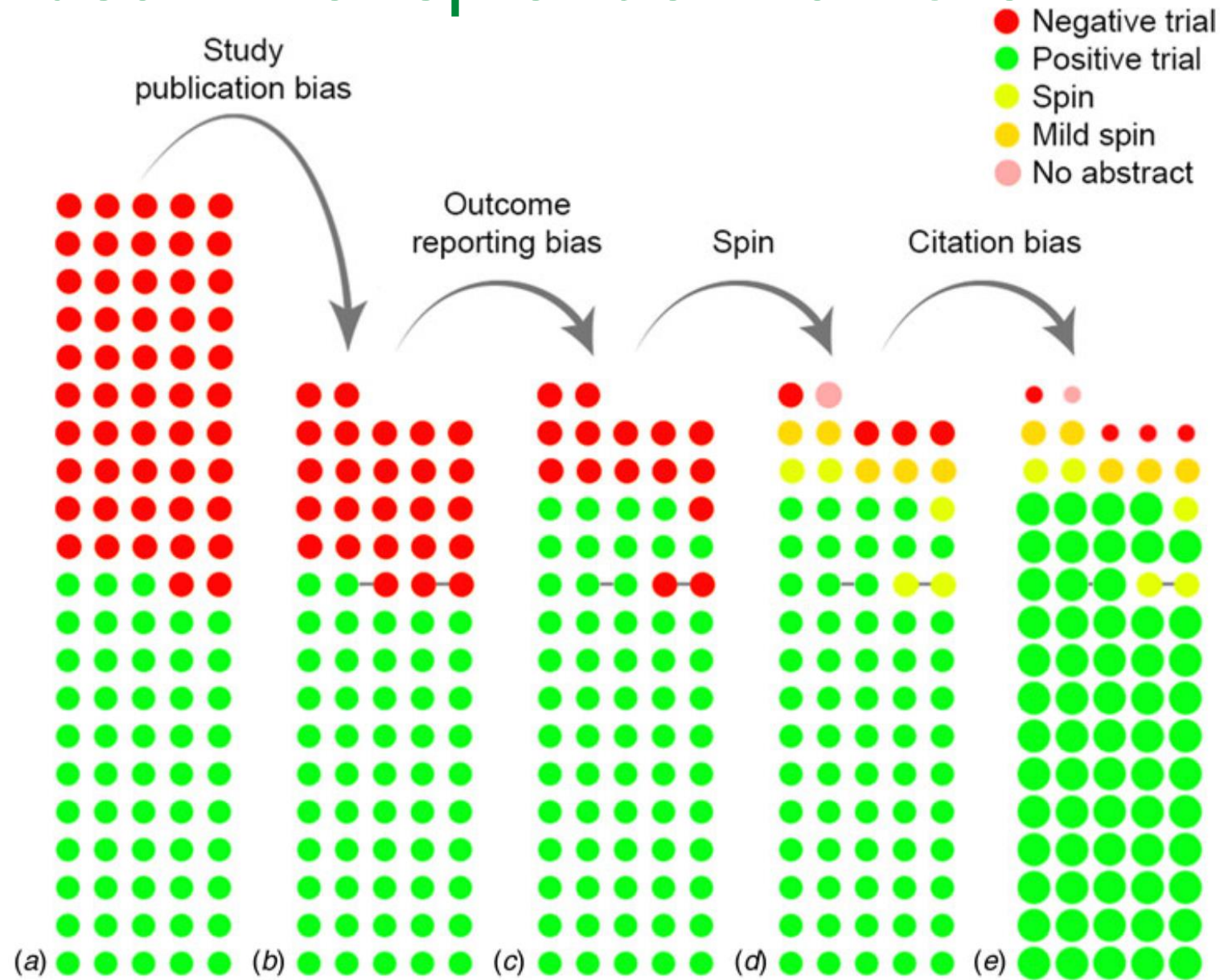
29. Juli 2023

QRP In concert



Zusammenspiel der Faktoren

Vries et al in Press





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

**We
can
DO
BETTER**

Open Science practices targeting collective action problem

- Pre-registration

Open Science practices targeting collective action problem

- Pre-registration



Open Science practices targeting collective action problem

- Pre-registration

**How does social media impact knowledge, polarization, and political participation?
A randomized field experiment on the effects of algorithmic ranking and virality on
Facebook and Instagram
PRE-ANALYSIS PLAN**

This text represents the intentions of the research team at the time it is filed. We may deviate from this plan if unexpected issues arise, but we will report a “populated PAP” if we materially deviate from it (Duflo et al. 2020).

Abstract

This study will assess the impact of core Facebook and Instagram features associated with engagement and virality. It will focus on two elements that characterize modern social networks: (1) the ranking algorithm and (2) virality through sharing, using treatments that turn off each of these elements. The resulting manipulations will leave experimental subjects with feeds displaying posts in reverse chronological order or excluding reshares. We will investigate the causal effect of these changes on political attitudes and behavior, including possible substitution with other online and offline activity. We additionally characterize how these manipulations alter the user experience in terms of composition of the News Feed and time spent on/engagement with the platforms.

Research Questions/Hypotheses

To guard against drawing false inferences due to multiple comparisons, we will correct our p-values using the method described below. Feed (“f”) hypotheses and research questions will be tested using both the Facebook and Instagram reverse chronological feed treatment arms (T1a and T1b). Reshare (“r”) hypotheses and research questions will be tested through viewer-side reshare holdout on Facebook (T2). All comparisons are to the control group.

Primary hypotheses:

We prioritize predictions about the following outcomes. Additional detail on how the outcomes will be operationalized is under **Outcome Variables** below.

H1f: Reverse chronological feed will reduce polarization and negative perceptions of outgroups.

H1r: Reshares holdout will reduce polarization and negative perceptions of outgroups.

H2f: Reverse chronological feed will reduce campaign knowledge.

H2r: Reshares holdout will reduce campaign knowledge.

H3f: Reverse chronological feed will reduce online and offline political participation.

RQ1f: How will reverse chronological feed impact the type of content people see on their feeds?

RQ1r: How will reshares holdout impact the type of content people see on their feeds?

Secondary hypotheses:

Open Science practices targeting collective action problem

Main Outcomes (Survey Data: Platform Data and Tracking Data where noted)

Below we further detail the dependent variables underlying the outcomes specified in our hypotheses/research questions above.

To construct composite scales (where noted below), we use principal components analysis with varimax rotation from the candidate items in the list. (Following standard practice, we run PCA in the control group and predict the index for those in treatment using the same set of weights.) Items that do not clearly load on the same underlying dimension (i.e., the first principal component) will be analyzed separately.

H1: Polarization and negative perceptions of outgroups

We will rely on two sets of outcome measures: (1) affective polarization, and (2) issue polarization.

Affective polarization

Composite index of 3 measures below, each created as the difference between own party and other party [PID or PIDLEAN]. Those who lean toward neither party are eliminated from the analysis [PIDLEAN=Neither].

1. Difference in feeling thermometer scores between people who support the party the respondent prefers (0-100) and people who support the other party (0-100) [FT_[PEOPLEGROUPS]].
2. Difference in feeling thermometer scores between people running for office as the party the respondent prefers (0-100) and people running for office from the other party (0-100) [FT_[PEOPLEGROUPS]].

Open Science practices

- Pre-registration
- Registered Reports
 - Nature, Nature Human Behavior, Journal of Politics, Research and Politics, PLOS One, Journal of Experimental Political Science
 - Full List of Journals at cos.io/rr



Summary up to now

- Science as a social process
- The „scientific method“ and our responsibility
- Science as the best but an imperfect process for knowledge production
- Merton's Morns
- Incentives structures
- Solutions

Idiosyncratic modeling choices

Does **immigration** undermine public support for social policy?

D Brady, R Finnigan - American sociological review, 2014 - journals.sagepub.com

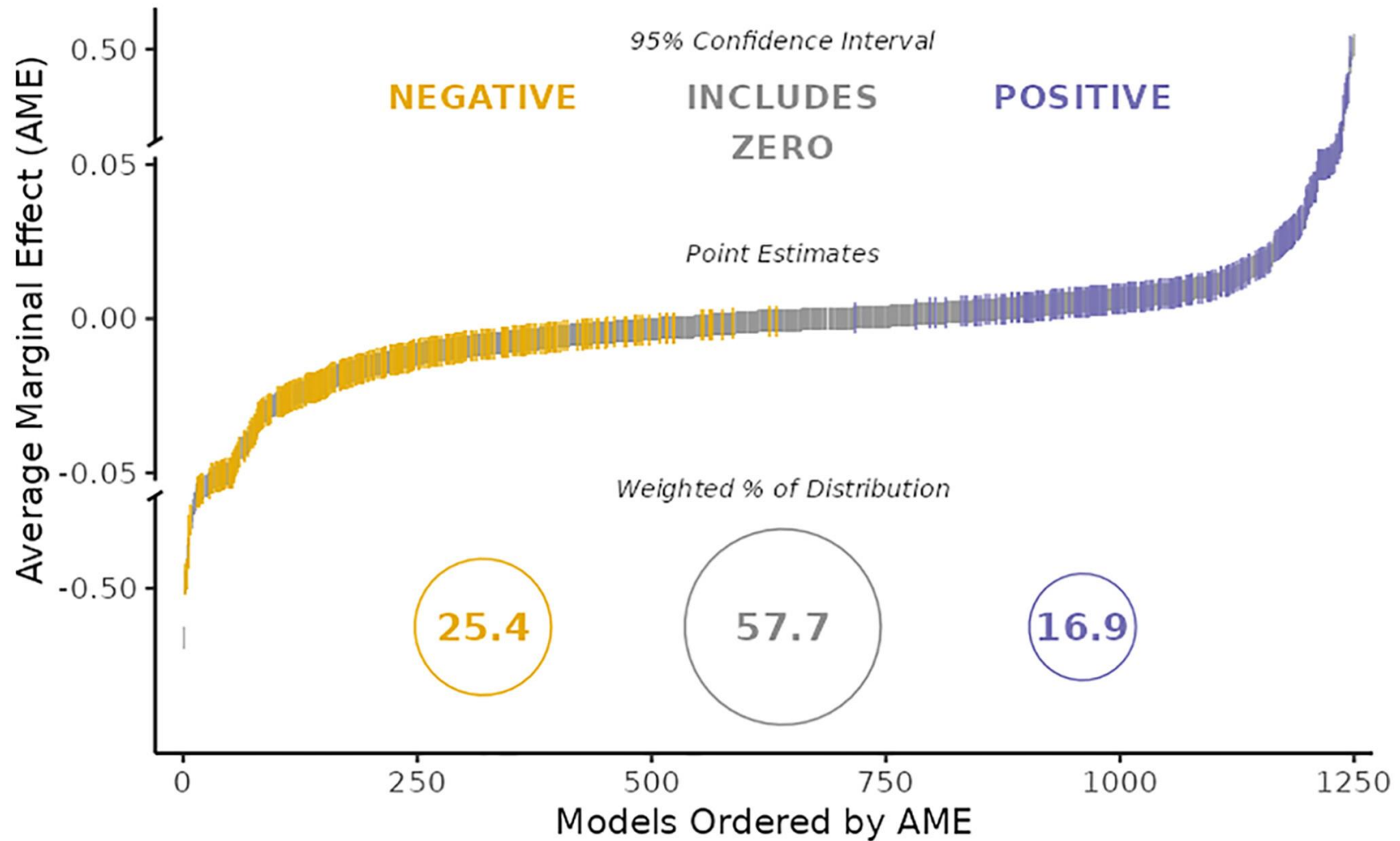
... -**immigrant** sentiments, many contend that **immigration** undermines public support for social policy. This study analyzes three measures of **immigration** and ... that **immigration** undermines ...

☆ Speichern Zitieren Zitiert von: 232 Ähnliche Artikel Alle 7 Versionen

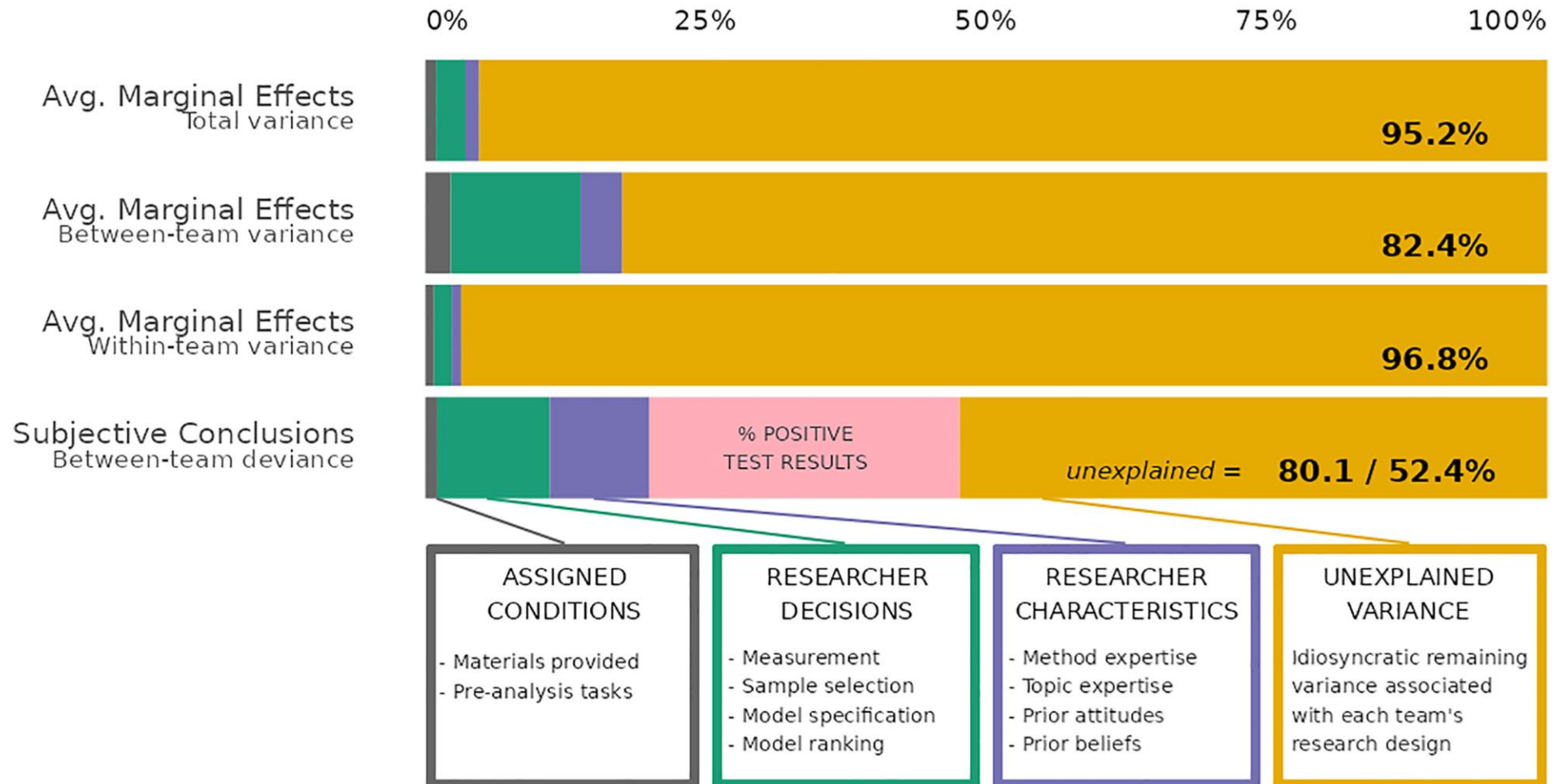
Abstract

There has been great interest in the relationship between immigration and the welfare state in recent years, and particularly since Alesina and Glaeser's (2004) influential work. Following literatures on solidarity and fractionalization, race in the U.S. welfare state, and anti-immigrant sentiments, many contend that immigration undermines public support for social policy. This study analyzes three measures of immigration and six welfare attitudes using 1996 and 2006 International Social Survey Program (ISSP) data for 17 affluent democracies. Based on multi-level and two-way fixed-effects models, our results mostly fail to support the generic hypothesis that immigration undermines public support for social policy. The percent foreign born, net migration, and the 10-year change in the percent foreign born all fail to have robust significant negative effects on welfare attitudes. There is evidence that the percent foreign born significantly undermines the welfare attitude that government "should provide a job for everyone who wants one." However, there is more robust evidence that net migration and change in percent foreign born have positive effects on welfare attitudes. We conclude that the compensation and chauvinism hypotheses provide greater potential for future research, and we critically consider other ways immigration could undermine the welfare state. Ultimately, this study demonstrates that factors other than immigration are far more important for public support of social policy.

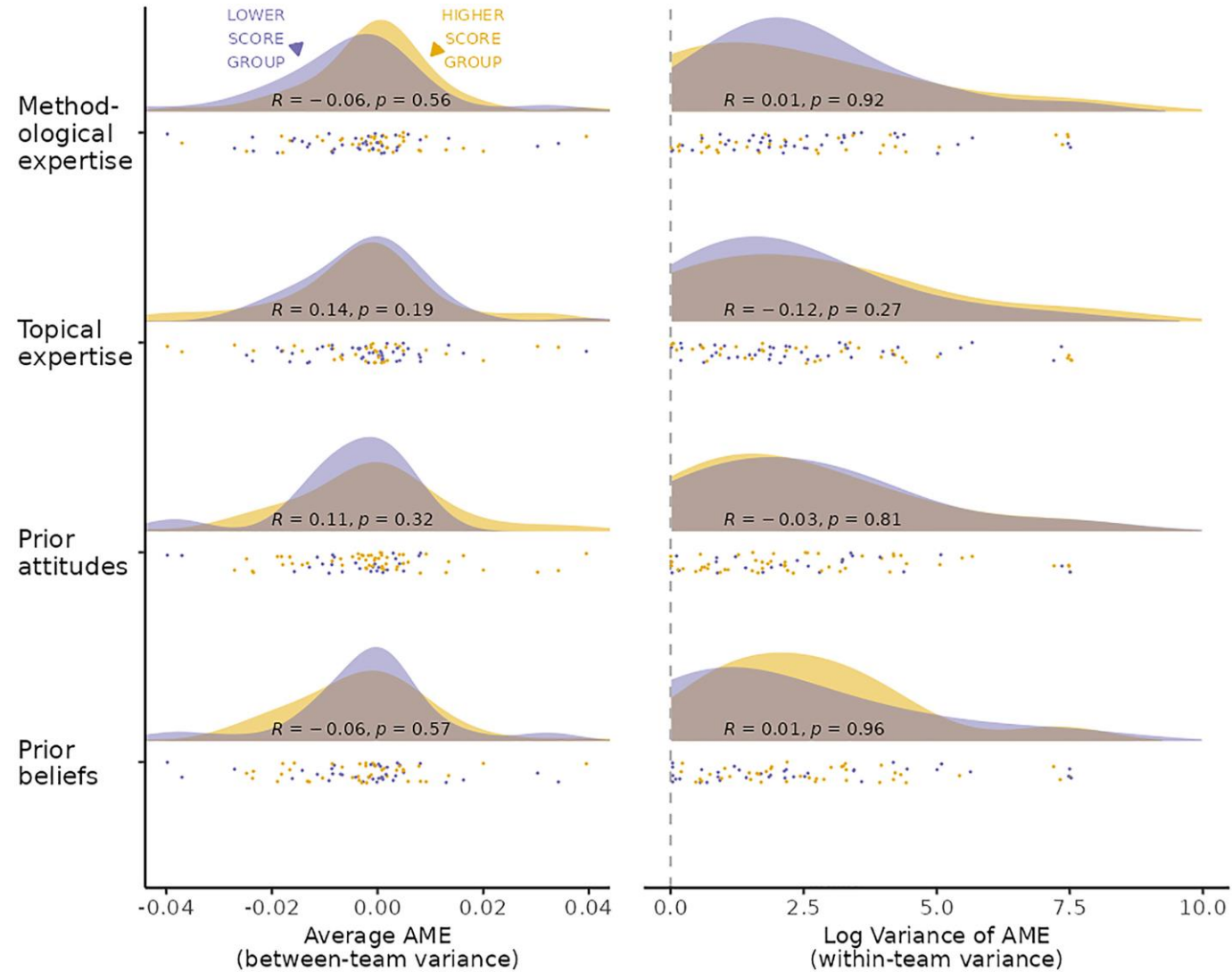
Heterogeneity in numerical results



Factors Explaining Variance in Results



Idiosyncrasy



Idiosyncratic modeling choices in CSS

Idiosyncrasy in CSS

- **Lots of data** (that give many options for handling)
- **Little theory** (to inform data-processing decisions)
- **Complex methods** (many options for handling, little understanding among users and reviewers)

Idiosyncrasy in text pre-processing

- Punctuation
- Numbers
- Lowercasing
- Stemming
- Stopword removal
- N-gram inclusion
- Infrequently used terms
- ...

Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It*

Matthew J. Denny[†]

Arthur Spirling[‡]

Abstract

Despite the popularity of unsupervised techniques for political science text-as-data research, the importance and implications of preprocessing decisions in this domain have received scant systematic attention. Yet, as we show, such decisions have profound effects on the results of real models for real data. We argue that substantive theory is typically too vague to be of use for feature selection, and that the supervised literature is not necessarily a helpful source of advice. To aid researchers working in unsupervised settings, we introduce a statistical procedure and software that examines the sensitivity of findings under alternate preprocessing regimes. This approach complements a researcher's substantive understanding of a problem by providing a characterization of the variability changes in preprocessing choices may induce when analyzing a particular dataset. In making scholars aware of the degree to which their results are likely to be sensitive to their preprocessing decisions, it aids replication efforts.

preText software available: github.com/matthewjdenny/preText

Idiosyncrasy in text pre-processing

Citation	Steps	Cites
Slapin and Proksch (2008)	P-S-L-N-W	427
Grimmer (2010)	L-P-S-I-W	258
Quinn et al. (2010)	P-L-S-I	275
Grimmer and King (2011)	L-P-S-I	109
Roberts et al. (2014)	P-L-S-W	117

Table 1: Preprocessing steps taken/suggested in recent notable papers that deal with unsupervised learning methods. The cite total is taken from Google Scholar at the time of writing. In the case of Slapin and Proksch (2008), we consulted their Wordfish manual (version 1.3). In the case of Roberts et al. (2014), the authors suggest further steps might be appropriate for a given application.

Model dependency in left-right scaling

$\text{Lab}_{1983} < \text{Lab}_{1987} < \text{Lab}_{1992} < \text{Lab}_{1997} < \text{Con}_{1992} < \text{Con}_{1997} < \text{Con}_{1987} < \text{Con}_{1983}$.

Specification	Most Left	Most Right
P-N-S-W-3-I	Lab_{1983}	Cons_{1983}
N-S-W-3	Lab_{1987}	Cons_{1987}
N-L-3	Lab_{1992}	Cons_{1987}
N-L-S	Lab_{1983}	Cons_{1992}

Table 3: Some example specifications which differ in terms of the manifestos they place on the (far) left and (far) right under the Wordfish model.

Model dependency in topic modeling

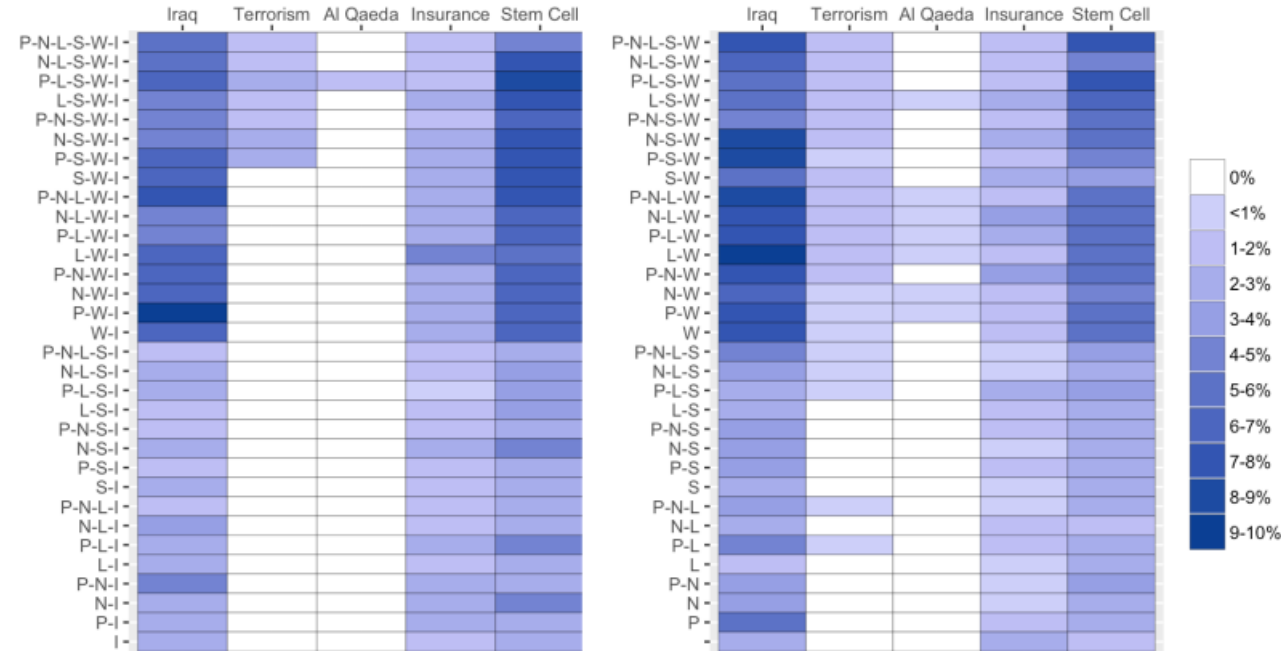


Figure 3: Plots depicting the percentage of topic top-20-terms which contain the stem of each of five keywords, for each of 64 preprocessing steps (thus excluding those which include trigrams). The number of topics for specifications fit to each of the 64 DFMs were determined through ten-fold cross validation, minimizing the model perplexity.



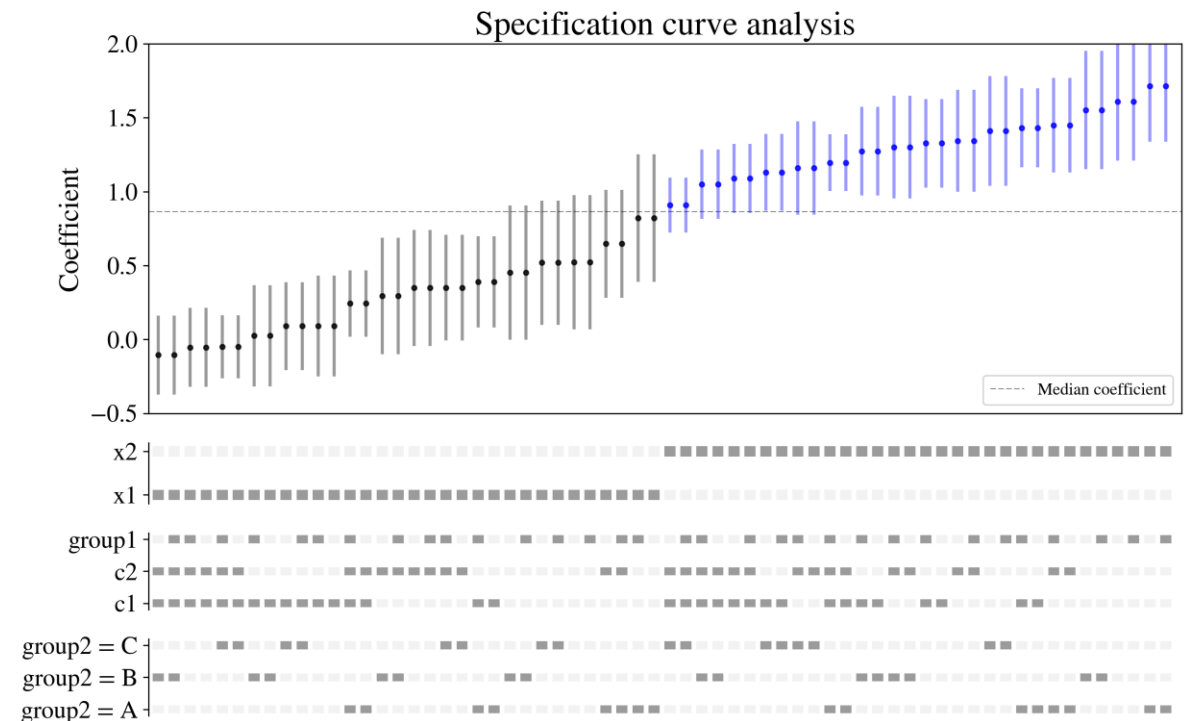
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

**We
can
DO
BETTER**

**Rule of thumb:
When multiple modeling decisions are equally plausible,
take all of them**

Solutions

- For text preprocessing: PreText
- General: Specification curve / Boba / Multiverse
- General: Interactive Shiny Appendices





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Reproducibility

What is reproducibility

Table 1: two-by-two grid defining reproducibility and replicability.

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Based on Computational Reproducibility in Computational Social Science
by David Schoch, Chung-hong Chan, Claudia Wagner, Arnim Bleier
<https://arxiv.org/abs/2307.01918>

Reproducibility, Definition

“A result is reproducible when the **same analysis** steps performed on the **same dataset** consistently produces the **same answer**.”

The Turing Way Community 2022

State of reproducibility

- Among 250 randomly selected social science studies, **raw data** were **available in only 7 percent of cases** and analysis scripts in only **1 percent** (Hardwicke et al. 2020),
- Stockemer et al. (2018) showed that mechanistic **correspondence between data and findings** could **not be established in a quarter of the studies**, often because of insufficient data organization or methodological transparency
- In **five percent**, reproduction attempts produced **significantly different results** than those reported in the original study (Stockemer et al. 2018)
- **Reinhart / Rogoff** (2010) incident

State of data sharing

- only 26% of data could still be obtained for six year old papers
- (Vines et al. 2014)
- Among articles stating that data was available upon request, only 17% shared data upon request (Hussey 2023)

Data is not available upon request

Ian Hussey

Many journals now require data sharing and require articles to include a Data Availability Statement. However, several studies over the past two decades have shown that promissory notes about data sharing are rarely abided by and that data is generally not available upon request. This has negative consequences for many essential aspects of scientific knowledge production, including independent verification of results, efficient secondary use of data, and knowledge synthesis. Here, I assessed the prevalence of data sharing upon request in articles employing the Implicit Relational Assessment Procedure published within the last 5 years. Of 52 articles, 42% contained a Data Availability Statement, most of which stated that data was available upon request. This rose from 0% in 2018 to 100% in 2022. Only 27% of articles' authors actually shared data. Among articles stating that data was available upon request, only 17% shared data upon request. The presence of Data Availability Statements was not associated with higher rates of data sharing ($p = .35$). Results replicate those found elsewhere: data is generally not available upon request, and promissory Data Availability Statements are typically not adhered to. Issues, causes, and implications are considered.



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

**We
can
DO
BETTER**

Reproducible by whom?

Table 2: Execution Matrix. This table shows the tiers of computational reproducibility based on the agent that conducts the reproducibility test and the computational environments that the agent uses for the test.

Agent \ Environment	Environment		
	Local	Restrictive	Nonrestrictive
Author(s) only	Unchecked	Unchecked	1°CR
Trusted third-parties	Not accessible, unchecked	2°CR	3°CR
Everyone	Not accessible, unchecked	Not accessible, unchecked	3°CR

Open Science practices

Code		sharable	Data not sharable
		Follow best practices Don't publish	Deposit with trusted 3rd party Don't publish
	sharable not sharable		

- Data sharing
 - Open Science Framework
 - (Github is not meant for long-term storage)
- Version control (This is what Github is for)
- Package management
 - Groundhog (For simpler projects)
 - Renv (For advanced projects)
- Reproducibility and computational environments
 - CodeOcean
 - Docker, see package containerit <https://joss.theoj.org/papers/10.21105/joss.01603>