



Quantitative Empirical Accounting Research

**Doctoral Seminar
07.-14.09.2020**



SFB/Transregio 266
**ACCOUNTING FOR
TRANSPARENCY**

Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - The open science workflow
 - Software: Github, R, Python, Stata
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation

Wednesday, 09.09.2020

- Theory
 - Equity-related incentives
 - Debt-related incentives
 - The role of the auditing process
- Research design
 - From the research question to the research setting
 - Identification strategies
- Paper: Gassen and Muhn (2018)
- Execution
 - Data wrangling
 - Exploration
 - Modelling and testing
- Class project
 - Exploratory data analysis
 - Discussion of potential research questions



Friday, 11.09.2020

- Execution
 - Assessing robustness
 - Writing a paper and presenting your work
 - Getting yourself published
- Paper: Noh, So and Verdi (2020)
- Workshop
 - Proposal presentations

Monday, 14.09.2020

- Class project
 - Result presentation and discussion
- Workshop
 - Proposal presentations
- Wrap-up

Structure of the seminar

The role of quantitative empirical research in social science

Descriptive studies

- Construct validity
- Exploratory data analysis
- Data visualization

Positive studies

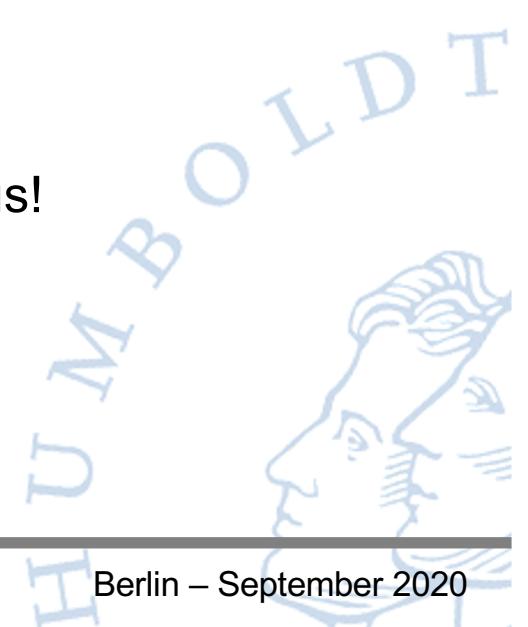
- Theory-based predictions
- Effect size estimation and power analysis
- Causal inference ...
 - when selection is based on observables
 - when selection based on unobservables

Execution

- Establishing a reproducible and collaborative workflow
- Identifying suitable data sources
- Tidying data
- Constructing samples
 - Variable definitions
 - Merging
- Presenting the analysis
- Assessing robustness

Some Guidelines

- This seminar is open and free form
- I aim at discussion-based learning not at frontal teaching
- Some stuff by me, presentations by you, interactive participation
- Feedback is crucial and will be given directly
- Especially the “execution” bit is mostly Q&A
- Most importantly: This should be rewarding and fun for all of us!



Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - Software: Github, R, Python, Stata
 - The open science workflow
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation

Theory: A supposedly simple question

What do you believe is going on?

Three seminal topics (and papers)

- Unraveling:

Milgrom (1981, BellJE): Good News and Bad News: Representation Theorems and Applications

- Earnings management:

Fischer and Verrecchia (2000, TAR): Reporting Bias

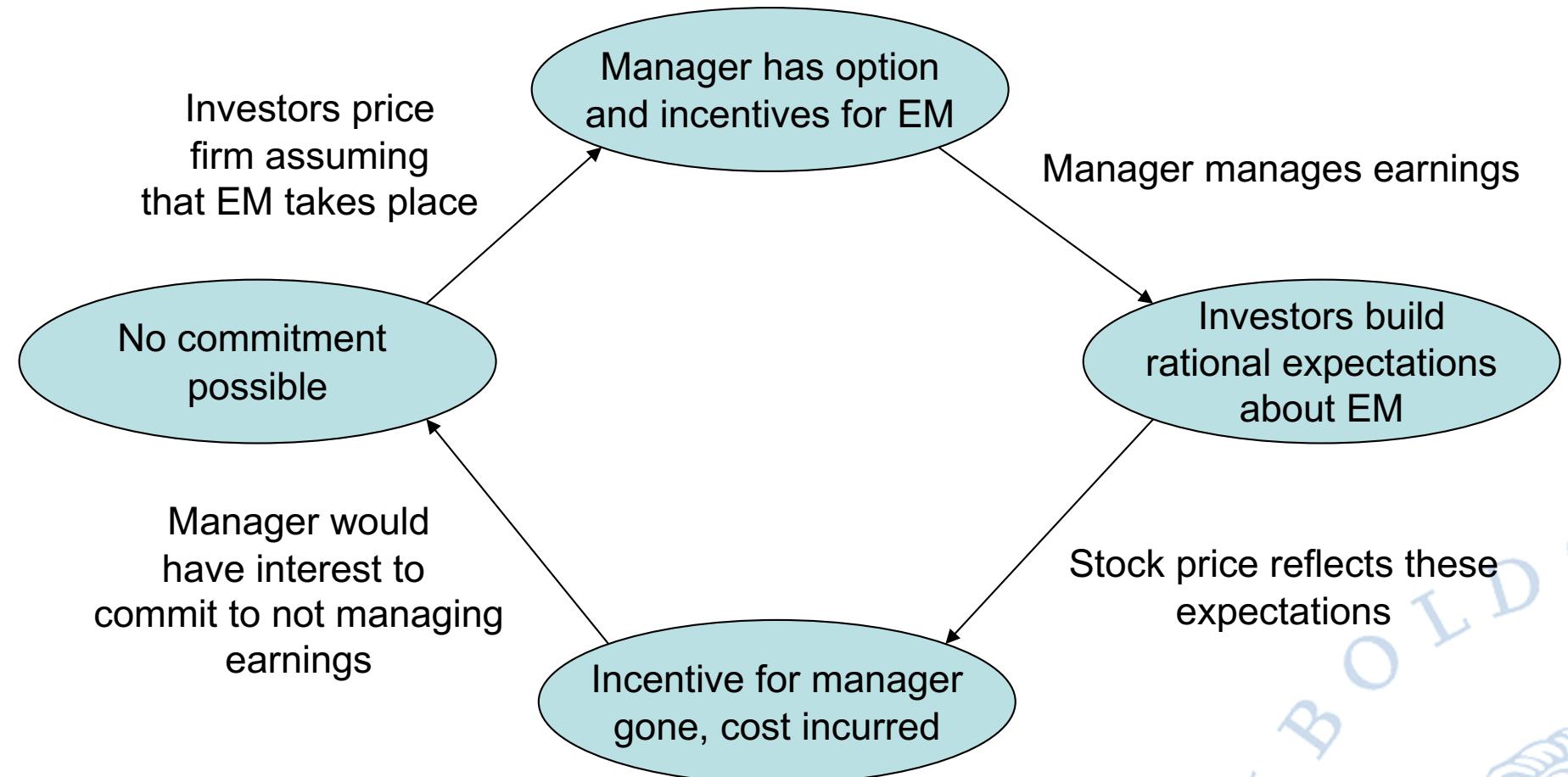
- Price discovery on capital markets:

Grossman and Stiglitz (1980, AER): On the Impossibility of Informationally Efficient Markets

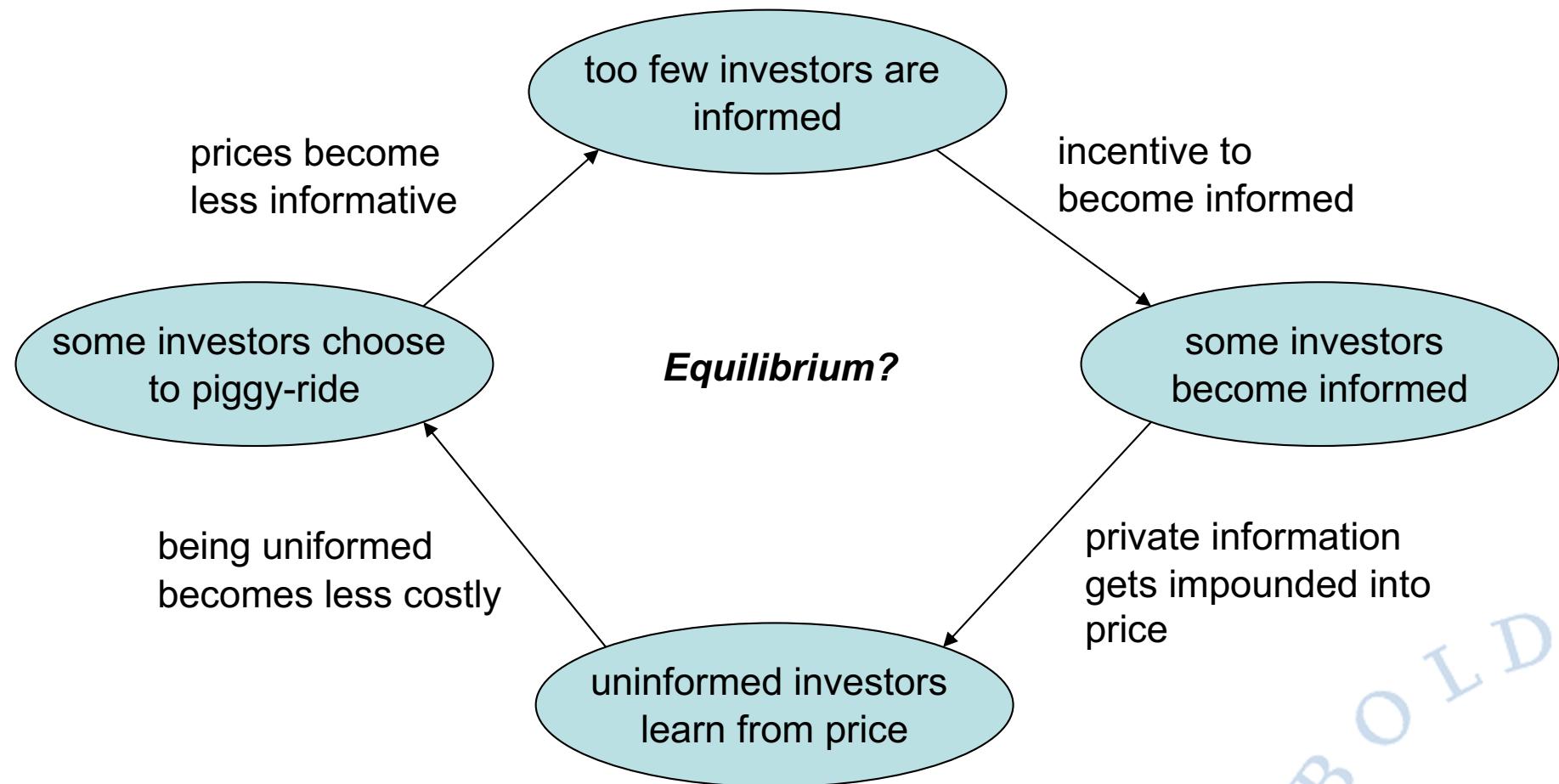
Unraveling

- Setup:
 - Group of firms (run by managers)
 - Investors build expectations on firm value
 - Value of firms can be ranked
 - Managers are informed about their firm's value
 - Everybody knows the distribution of firm values
- Assumptions:
 - Managers prefer higher stock price
 - Managers can truthfully disclose firm value at zero cost
 - Investors build rational expectations, are sceptical in doubt
- Outcome:
 - Every manager discloses (manager of firm with lowest firm value is indifferent)

Earnings management with rational expectations



Price discovery on capital markets



Grossman and Stiglitz (1980, p. 404):

“Efficient Markets theorists seem to be aware that costless information is a *sufficient* condition for prices to fully reflect all available information [...]; they are not aware that it is a *necessary* condition.”

Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - The open science workflow
 - Software: Github, R, Python, Stata
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation

Qualitative versus quantitative Research



Research approaches in accounting

- **Descriptive Studies**
 - *What is accounting like?*
 - Explanation and description of accounting
- **Positive Studies**
 - *Why is accounting like that?*
 - Development and testing of theories
 - Identify cause-and-effect relationship
- **Normative Studies**
 - *How should accounting be?*
 - Expresses normative (value) judgements
 - Uses theory, policy guidance likely to be backed up by causal evidence
 - Offers tools to improve object of study (rules, standards, ...)

Financial accounting research that is useful for reality...

1. Provide descriptive evidence on preparation and usage of financial reporting information
2. Develop positive theories about determinants and consequences of financial reporting information usage
3. Test these theories in suitable settings
4. Use gathered evidence to provide normative guidance on relevant financial reporting issues

See: Cascino et al. (AiE 2014)

The ideal situation: The classical experiment

Treatment D → Effect Y
(truly random)



ORIGINAL ARTICLE

Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game

S Kühn¹, T Gleich², RC Lorenz^{2,3}, U Lindenberger¹ and J Gallinat²

Video gaming is a highly pervasive activity, providing a multitude of complex cognitive and motor demands. Gaming can be seen as an intense training of several skills. Associated cerebral structural plasticity induced has not been investigated so far. Comparing a control with a video gaming training group that was trained for 2 months for at least 30 min per day with a platformer game, we found significant gray matter (GM) increase in right hippocampal formation (HC), right dorsolateral prefrontal cortex (DLPFC) and bilateral cerebellum in the training group. The HC increase correlated with changes from egocentric to allocentric navigation strategy. GM increases in HC and DLPFC correlated with participants' desire for video gaming, evidence suggesting a predictive role of desire in volume change. Video game training augments GM in brain areas crucial for spatial navigation, strategic planning, working memory and motor performance going along with evidence for behavioral changes of navigation strategy. The presented video game training could therefore be used to counteract known risk factors for mental disease such as smaller hippocampus and prefrontal cortex volume in, for example, post-traumatic stress disorder, schizophrenia and neurodegenerative disease.

Molecular Psychiatry (2014) **19**, 265–271; doi:10.1038/mp.2013.120; published online 29 October 2013

Keywords: longitudinal; plasticity; training; video gaming; voxel-based morphometry

A common problem...

RESEARCH WATCH

Smoking found to affect the brain

By Tannith Cattermole

09:19 December 10, 2010

6 Comments 



T
D
V
K
A
M

Reduced Thickness of Medial Orbitofrontal Cortex in Smokers

Simone Kühn, Florian Schubert, and Jürgen Gallinat

Background: Structural deficiencies within the prefrontal cortex might be related to drug-taking behavior that prevails in smokers. Cortical thickness has been found to be a structural modulator of cerebral function and cognition and a subtle correlate of mental disorders. However, to date an analysis of cortical thickness in smokers compared with never-smokers has not been undertaken.

Methods: We acquired high-resolution magnetic resonance imaging scans from 22 smokers and 21 never-smokers and used FreeSurfer to model the gray-white and pial surfaces for each individual cortex to compute the distance between these surfaces to obtain a measure of cortical thickness. The main cortical folds were aligned across individuals with FreeSurfer's surface-based averaging technique to compare whole brain differences in cortical thickness between smokers and never-smokers.

Results: Relative to never-smokers, smokers showed greater cortical thinning in the left medial orbitofrontal cortex (mOFC). Cortical thickness measures extracted from mOFC correlated negatively with the amount of cigarettes consumed/day and the magnitude of lifetime exposure to tobacco smoke.

Conclusions: The brains of smokers are structurally different from those of never-smokers in a dose-dependent manner. The cortical thinning in mOFC in smokers relative to never-smokers might imply dysfunctions of the brain's reward, impulse control, and decision-making circuits. Related behavioral correlates are suggested to be relevant for smoking initiation and maintenance.

0006-3223/\$36.00

doi:10.1016/j.biopsych.2010.08.004

BIOL PSYCHIATRY 2010;68:1061–1065

© 2010 Society of Biological Psychiatry

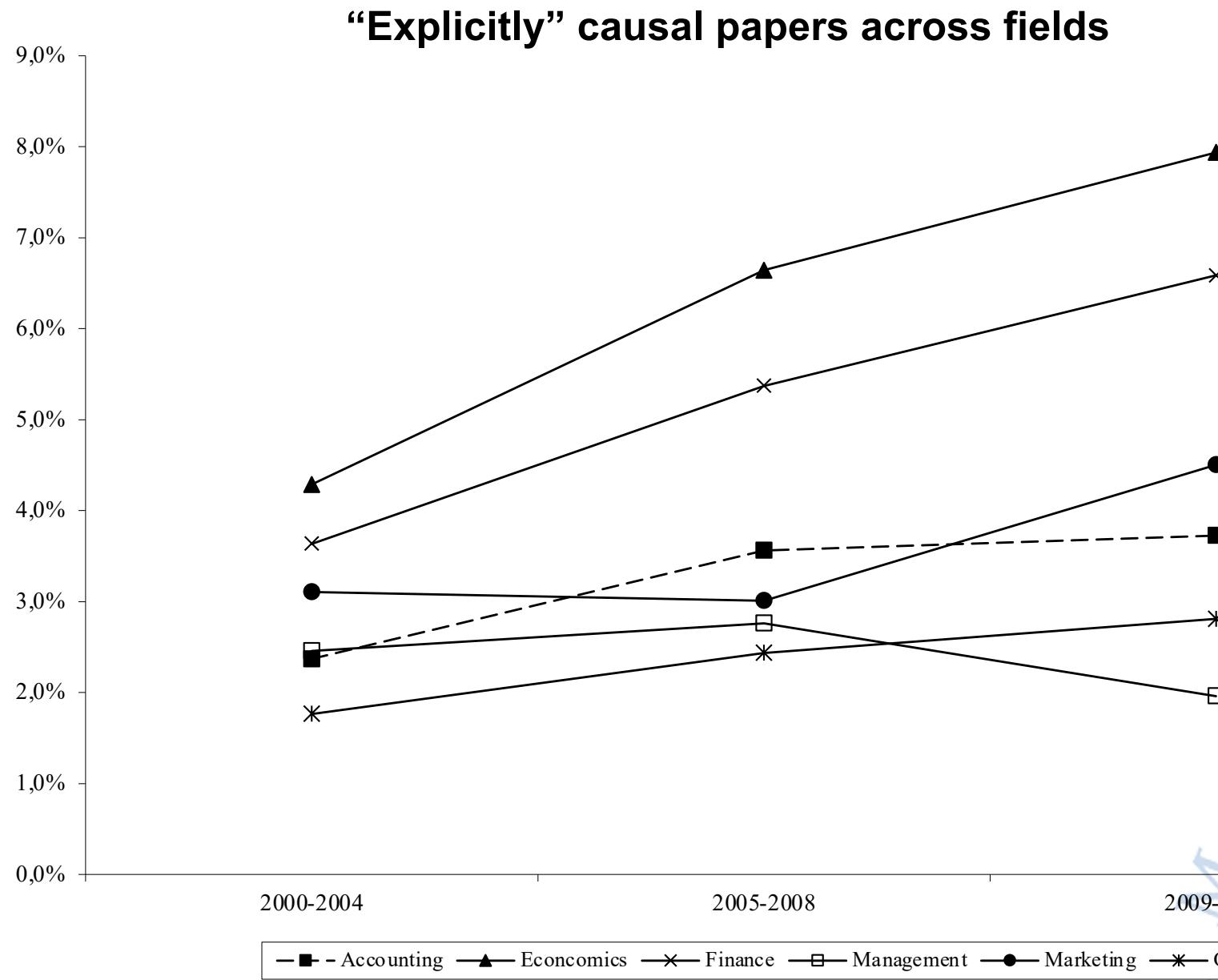
The core problem: unobservable counterfactuals

Treatment D	Selected to get no treatment ($Z=0$)	Selected to get treatment ($Z=1$)
0	Outcome $Y=10$	Outcome $Y=?$
1	Outcome $Y=?$	Outcome $Y=15$

A naive treatment estimator would be 5.

How do you calculate

- The average treatment effect (ATE),
- The average treatment effect of the treated (ATT)
- The average treatment effect of the control (ATC)?



Gassen (2014, AOS)

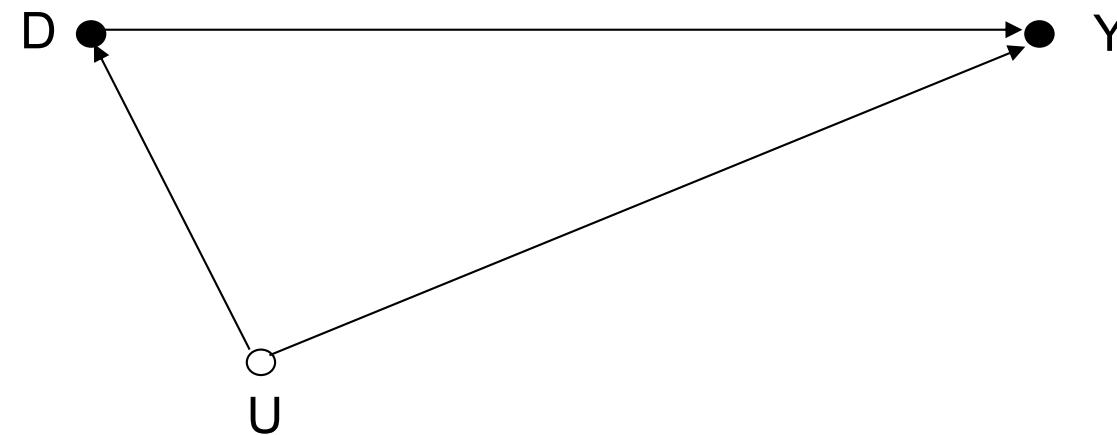
How papers “become causal”: Finance versus Accounting

Nature of Causal Focus	Financial Accounting	Finance
Motivation for robustness test	28 (68 %)	25 (29 %)
Motivation for statistical inference strategy	10 (24 %)	30 (35 %)
Motivation for research setting	3 (7 %)	30 (35 %)
Total	41	85

Gassen (2014, AOS)

Visualizing research designs

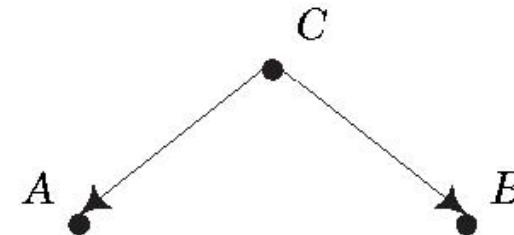
- Research designs can be analyzed by path models, in particular by directed acyclic graphs (DAGs)



Links between variables



(a) Mediation



(b) Mutual dependence

What effect does conditioning on C have on the observed relation between A and B in cases (b) and (c)?

Morgan and Winship, 2007, page 65



(c) Mutual causation

Figure 3.3: Basic patterns of causal relationships for three variables.

The effect of conditioning on a collider variable

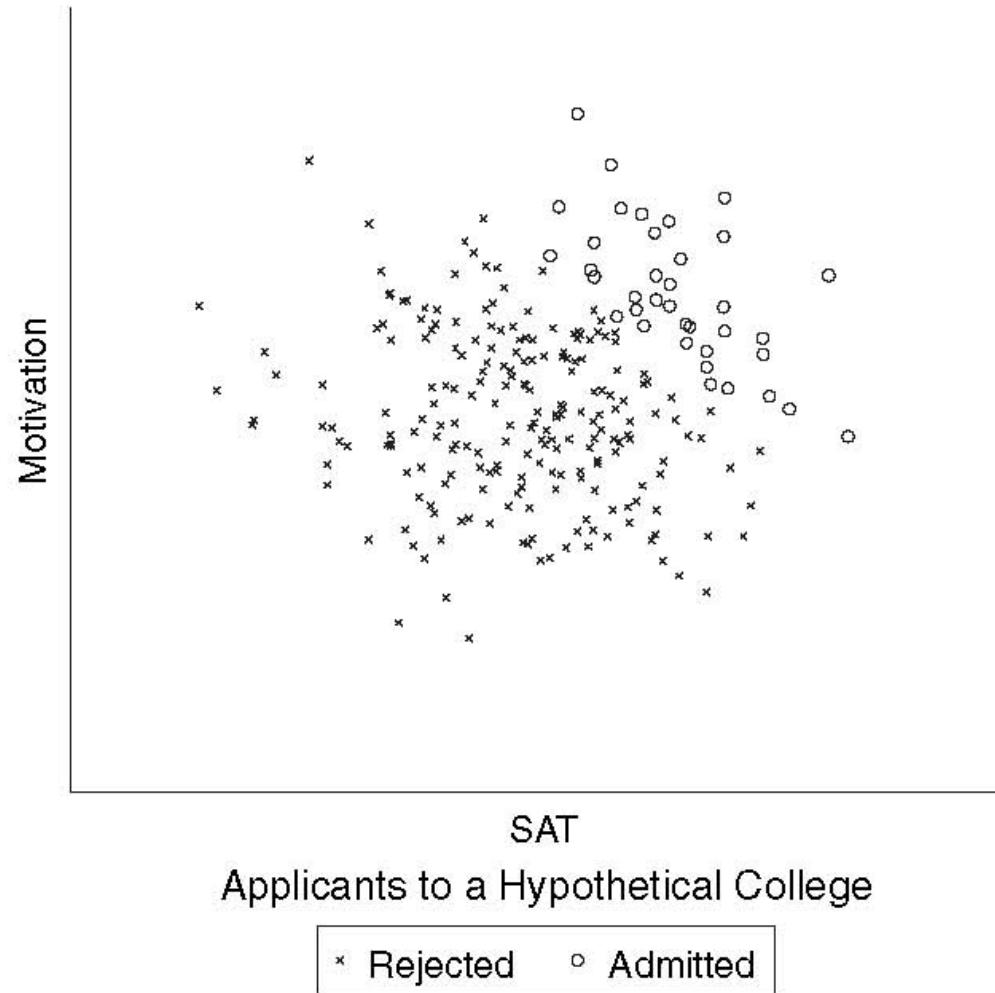
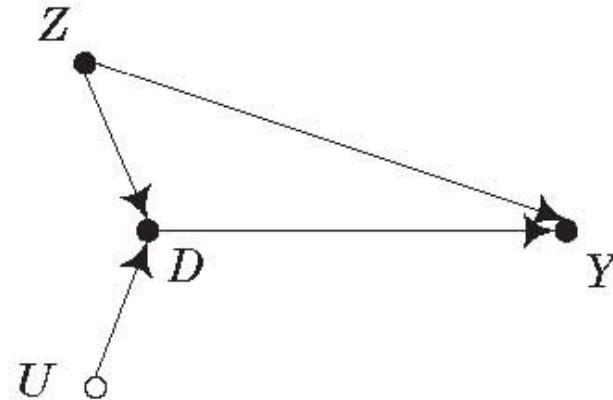


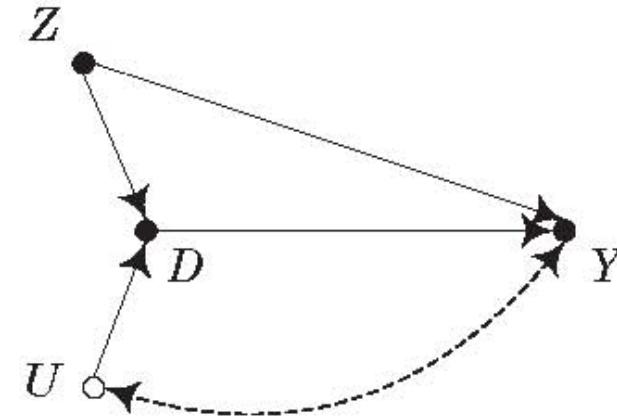
Figure 3.4: Simulation of conditional dependence within values of a collider variable.

Morgan and Winship, 2007, page 67

The problem of endogeneity



(a) Selection on the observables



(b) Selection on the unobservables

Figure 3.9: Causal diagrams for the terminology from econometric modeling of treatment selection.

Morgan and Winship, 2007, page 81

Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - The open science workflow
 - Software: Github, R, Python, Stata
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation

Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - The open science workflow
 - Software: Github, R, Python, Stata
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation

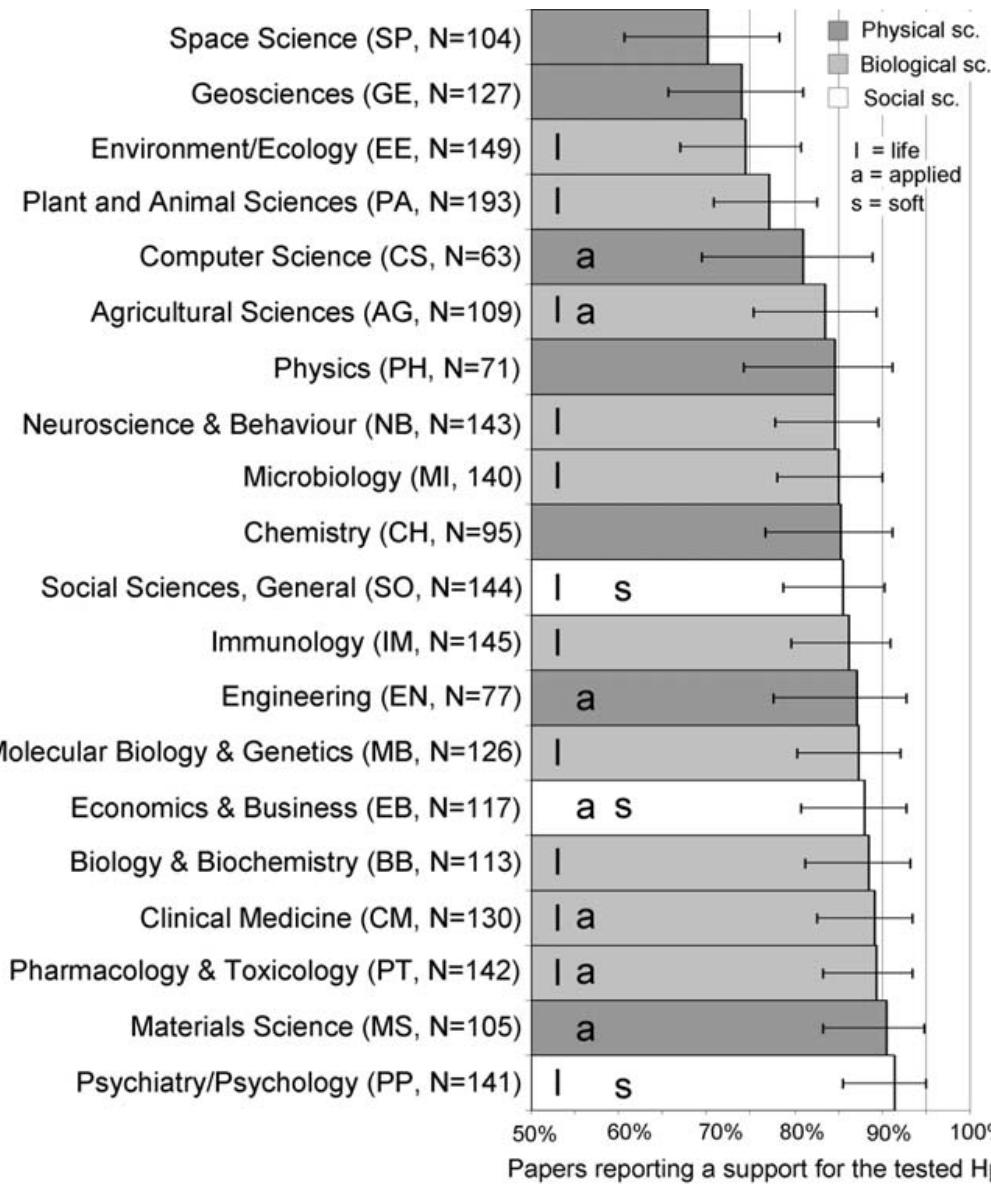
Open Science

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

— FOSTER, <https://www.fosteropenscience.eu/>

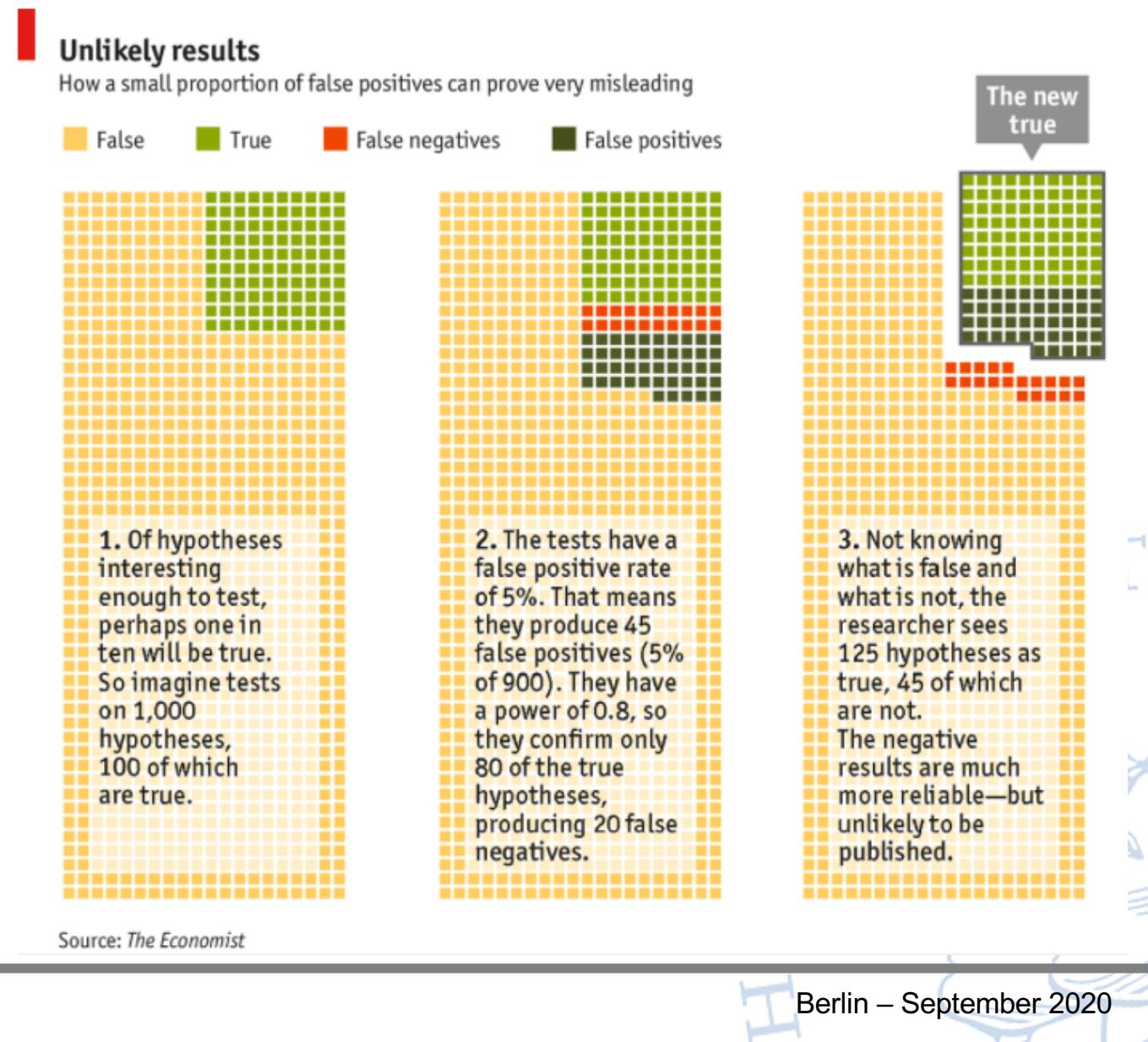


Do we have a problem?



Fanelli D (2010) “Positive” Results Increase Down the Hierarchy of the Sciences. PLoS ONE 5(4): e10068

False positives



P-hacking: A simplified mind experiment

- Assume that you have a project with
 - three different potential proxies for your dependent variable
 - of which one can be measured in two different ways
 - and one main experimental variable, alternatively proxied for by two different measures
 - and a set of mandatory control variables plus
 - five additional control variables that you could or could not include
- Playing the devil's advocate, assuming H0 (no significant relation of the experimental variable with the dependent construct) and independence of the different measures:
- Roughly how many significant results (two-sided below 5 %) can you expect to produce?

$$4 (y) * 2 (x) * 2^5 (z) = 256 \text{ regressions} * 0.05\% \approx 13$$

Replications...

The screenshot shows the homepage of the journal **Science**, published by AAAS. The main navigation menu includes Home, News, Journals, Topics, and Careers. Below the menu, there are links to various sub-journals: Science, Science Advances, Science Immunology, Science Robotics, Science Signaling, and Science Translational Medicine. The current page is Science.

SHARE RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration*†

+ See all authors and affiliations

Science 28 Aug 2015:
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

19

Article Figures & Data Info & Metrics eLetters PDF

You are currently viewing the abstract.

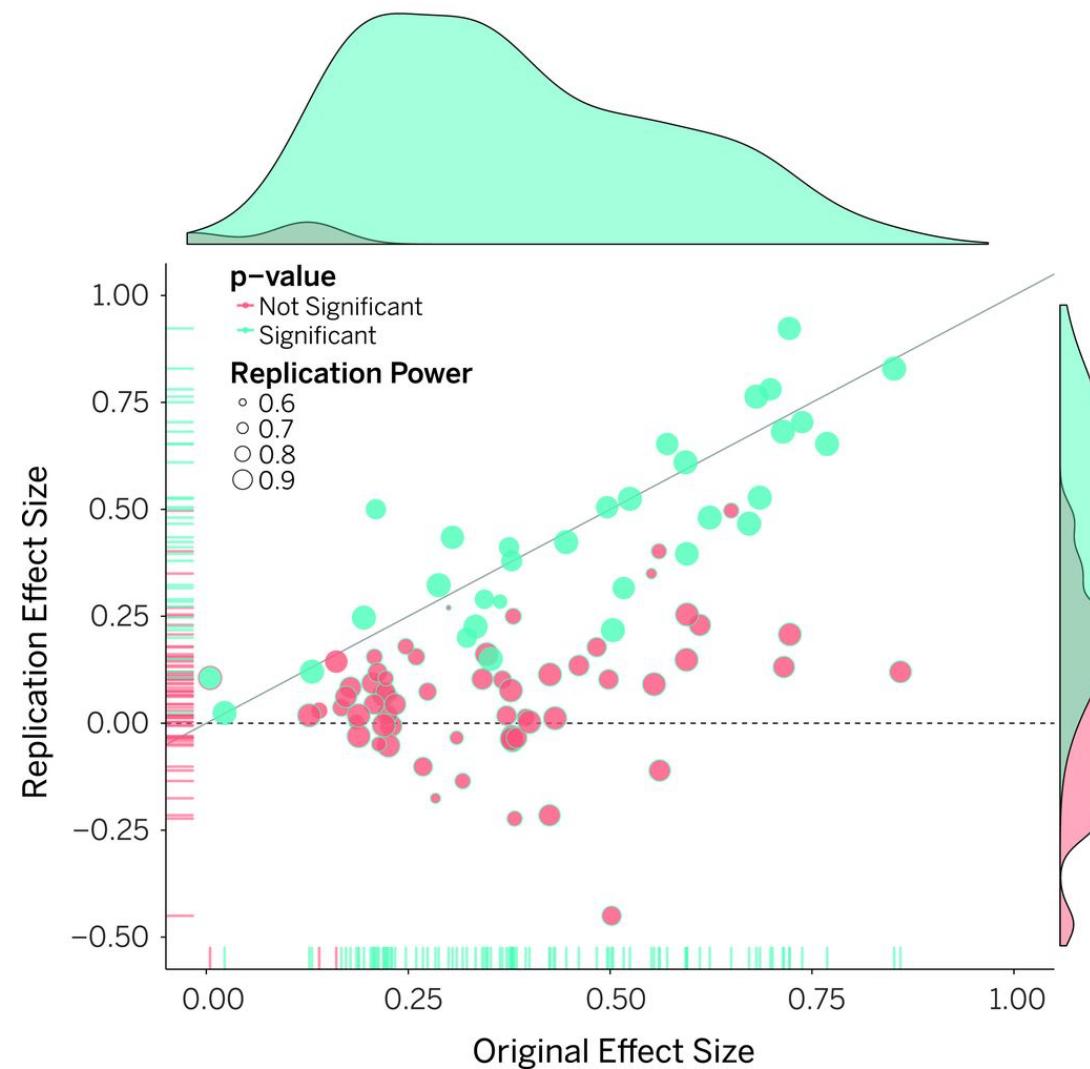
View Full Text ▶

Empirically analyzing empirical evidence

One of the central goals in any scientific endeavor is to understand causality. Experiments that seek to demonstrate a cause/effect relation most often manipulate the postulated causal factor. Aarts *et al.* describe the replication of 100 experiments reported in papers published in 2008 in three high-ranking psychology journals. Assessing whether the replication and the original experiment yielded the same result according to several criteria, they find that about one-third to one-half of the original findings were also observed in the replication study.

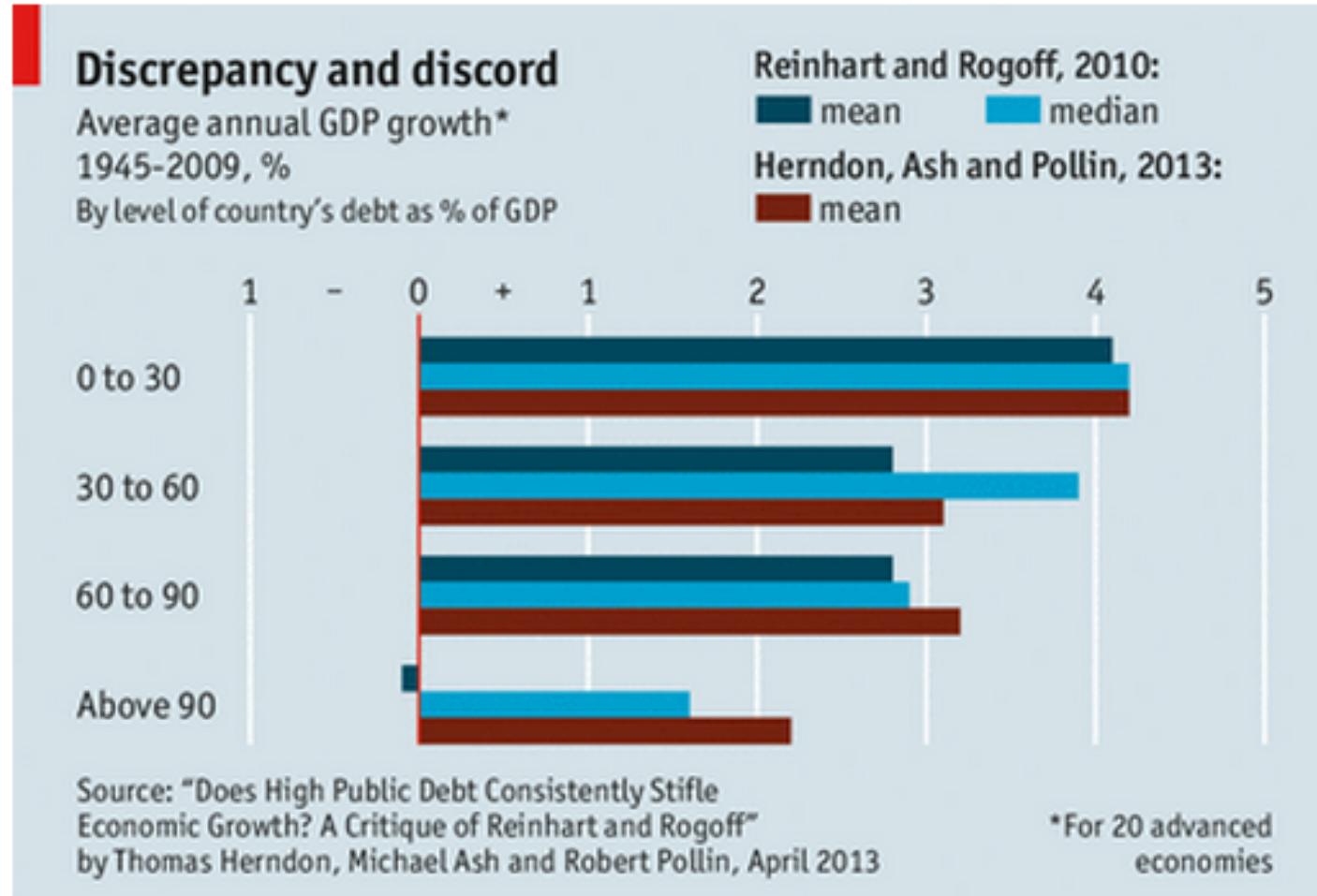
Science, this issue [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)

...tend to show discomforting results



Open Science Collaboration Science 2015;349:aac4716

Another and maybe even more important issue...



Source: The Economist, The 90% question, April 23, 2013

Things to consider when designing (open science) projects

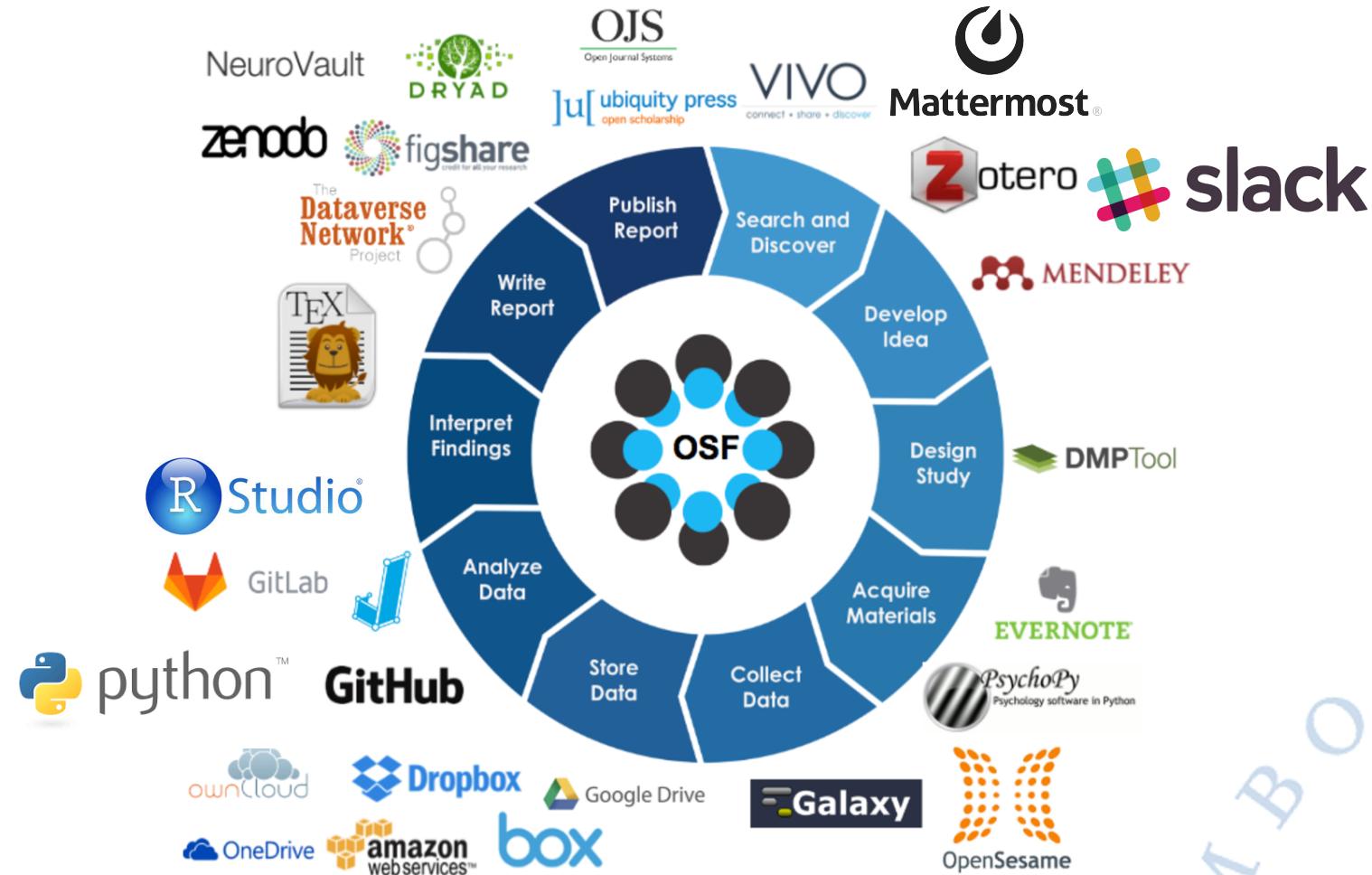
- Can one build an environment that allows reproducing my analysis with reasonable resources?
- If I am using commercially licensed Software: Is this software essential for my analysis? (think about Stata)
- Is the data that I use publicly obtainable and have I documented where to get it from?
- If the data is not publicly available: Is it essential for my analysis?
- Is the code that I use relying on some idiosyncrasies of my development environment (paths, support software)?
- Is my code readable?
- Have I tried to replicate my analysis in different environments?

The scientific workflow...



<https://cos.io/our-products/open-science-framework/>

... and its support by software



<https://cos.io/our-products/open-science-framework/>,
with extensions

An open source-based workflow

- Revision control system: Git/Github
- Relational database: e.g., MySQL
- General purpose programming language: Python
- Statistical programming language: R
- Typesetting program for papers and presentations: LaTeX



My take on the big 4 statistical programming languages

- Julia:
 - Pros: Fast, open source, expert user base
 - Cons: Small user base, few packages
- Python:
 - Pros: General purpose, open source, relatively easy to learn, many machine learning packages, large user base
 - Cons: Packaging system, statistic packages have limited interoperability, object orientation feels alien when working with data
- R:
 - Pros: Focused on data science, open source, packaging system, interoperability, graphics system
 - Cons: Not really easy to learn, tidyverse helps though
- Stata
 - Pros: Easy to learn, very broad user base in economics, most statistical methods are quickly implemented
 - Cons: Commercially licensed and closed source, inflexible programming environment

Monday, 07.09.2020

- Theory
 - The positive accounting research paradigm
 - Microeconomic foundations
 - Asset pricing foundations
- Research design
 - Descriptive versus causal studies
 - The counterfactual framework
- Paper: Breuer and Windisch (2019)
- Execution
 - The open science workflow
 - Software: Github, R, Python, Stata
 - Data sources and data constructs
- Class project
 - Introduction of setting and question
 - Data presentation