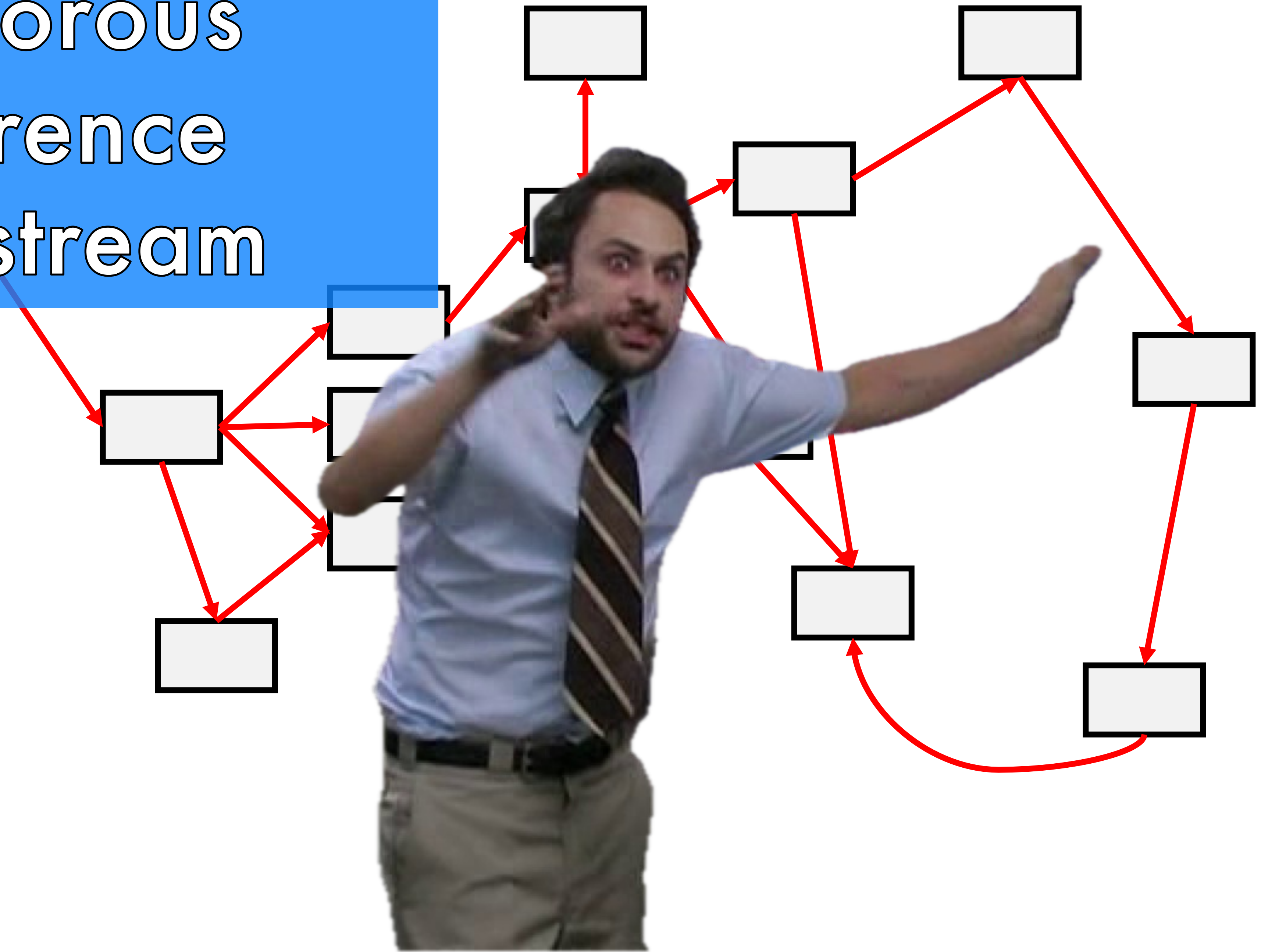
Slides: juliarohrer.com/resources

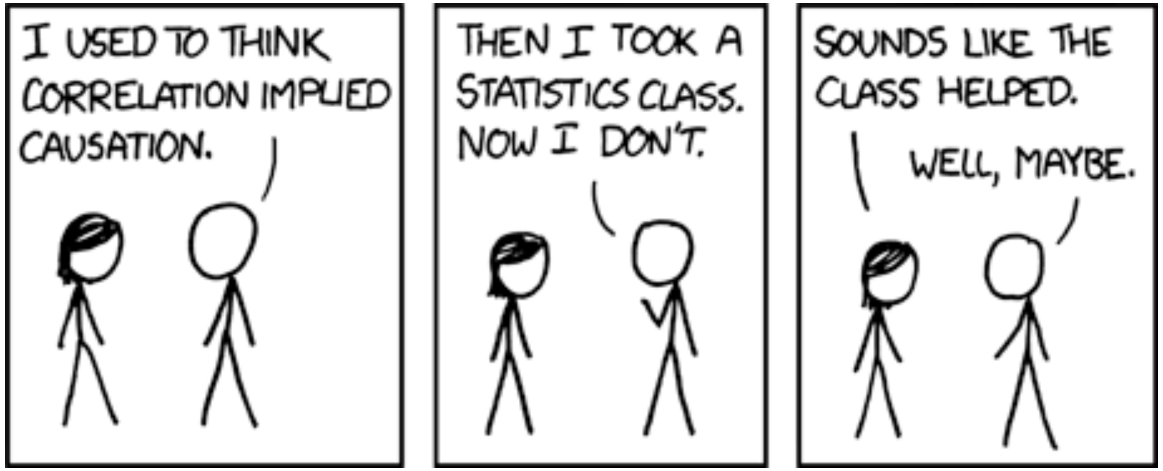
Making rigorous causal inference more mainstream



Julia M. Rohrer
Leipzig University
www.juliarohrer.com
@dingdingpeng.the100.ci
www.the100.ci

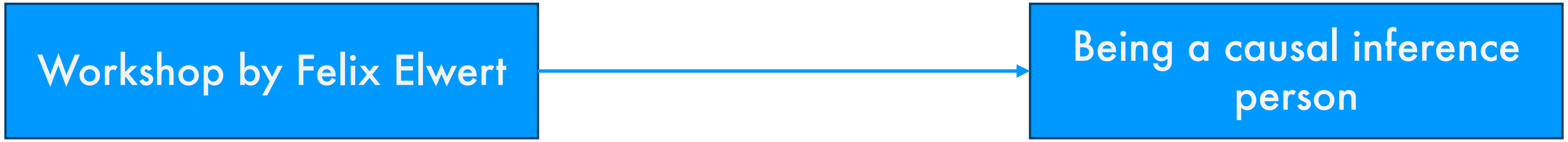
Slides: juliarohrer.com/resources

Causal Inference 101, according to psychology



- Option 1: Run an experiment
- Option 2: Give up
  - include covariates maybe?
  - or maybe use longitudinal data???

Comic: https://xkcd.com/552/


Workshop by Felix Elwert
Being a causal inference person

lides: juliarohrer.com/resources

Alexander Mass on Unsplash

Causal inference problems

- Researchers don‘t even realize or acknowledge that their research question requires causal reasoning
- Researchers have no idea how to connect their causal research questions to data and no idea which assumptions are involved
- Researchers draw causal inferences that rest on heroic assumptions

The non-experimentalists dilemma

1. virtually all interesting research questions concern causality
2. observational data are not admissible for causal inference
3. you work on something that cannot be experimentally randomized

1. + 2. + 3. = ???

[Grosz, Rohrer, & Thoemmes (2020)]

The non-experimental psych workaround

- introduction: relies on a causal reading of the literature
- methods & results: implicitly (but never explicitly) causal inference-y
  - X predicts Y…even after accounting for…
  - X is a risk factor for Y
  - longitudinal associations
- discussion: only makes sense in terms of causality
- **IMPORTANT:** add a paragraph that your study was only observational and no causal conclusions are warranted
  - future longitudinal or experimental studies will surely fix this problem

[Grosz, Rohrer, & Thoemmes (2020)]

Slides: juliarohrer.com/resources

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Income and emotional well-being: A conflict resolved

Matthew A. Killingsworth[a,1], Daniel Kahneman[b,2], and Barbara Mellers[a]

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received May 20, 2022; accepted November 2

Do larger incomes make people happier? Two authors of the present paper have published contradictory answers. Using dichotomous questions about the preceding day, [Kahneman and Deaton, Proc. Natl. Acad. Sci. U.S.A. 107, 16489–16493 (2010)] reported a flattening pattern: happiness increased steadily with log(income) up to a threshold and then plateaued. Using experience sampling with a continuous scale, [Killingsworth, Proc. Natl. Acad. Sci. U.S.A. 118, e2016976118 (2021)] reported a linear-log pattern in which average happiness rose consistently with log(income). We engaged in an adversarial collaboration to search for a coherent interpretation of both studies. A reanalysis of Killingsworth's experienced sampling data confirmed the flattening pattern only for the least happy people. Happiness increases steadily with log(income) among happier people, and even accelerates in the happiest group. Complementary nonlinearities contribute to the overall linear-log relationship. We then explain why Kahneman and Deaton overstated the flattening pattern and why Killingsworth failed to find it. We suggest that Kahneman and Deaton might have reached the correct conclusion if they had described their results in terms of unhappiness rather than happiness; their measures could not discriminate among degrees of happiness because of a ceiling effect. The authors of both studies failed to anticipate that increased income is associated with systematic changes in the shape of the happiness distribution. The mislabeling of the dependent variable and the incorrect assumption of homogeneity were consequences of practices that are standard in social science but should be questioned more often. We flag the benefits of adversarial collaboration.

Well-being | happiness | income | income satiation | experience sampling

Can money buy happiness? Two authors of this article have published contradictory claims about the relationship between emotional well-being and income. We later agreed that both studies produced valid results and that it was our responsibility to search for an interpretation that explains both findings. We engaged in an adversarial collaboration and

Si
Me
oft
(in
[Pr
16
ex
em
inc
so
$9
[Pr
e2
lin
an
sa
a j
ex
the
res
of
co
co
rel

The main finding of our reanalysis of MK's study is that the shape of the distribution of happiness changes—slightly, but systematically—as income rises. The same increases of income have different effects on the happy and on the unhappy regions of the distribution. In the low range of incomes, unhappy people gain more from increased income than happier people do. In other words, the bottom of the happiness distribution rises much faster than the top in that range of incomes. The trend is reversed for higher incomes, where very happy people gain much more from increased income than unhappy people do. The upper part of the happiness distribution rises with log(income) at an accelerated rate in that range, while the lower 20% is almost completely flat. The middle of the happiness distribution shows approximately linear gains in happiness with rising log(income). We use terms such as "increase" and "gain" for ease of exposition but, to be clear, we are simply describing cross-sectional associations between happiness and income (just as KD and MK did).

The results of this analysis violate a homogeneity assumption that is routinely made—and rarely checked—in the study of bivariate relationships. The assumption is more restrictive than the familiar condition of homoscedasticity. It holds if the conditional distributions of the predicted variable retain the same shape over the entire range of the predictor. In the present case, homogeneity requires the entire distribution of happiness to move in

LETTER | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Inappropriate causal assumptions underlie Killingsworth, Kahneman, and Mellers' conclusions

Julia M. Rohrer and Sebastian E. Wenz Authors Info & Affiliations

October 31, 2024 | 121 (46) e2313712121 | https://doi.org/10.1073/pnas.2313712121

VIEW THE ORIGINAL ARTICLE + THIS ARTICLE HAS A REPLY +

Killingswor

whether m

averages t

quantile re

log(income

log(income

Rohrer and Wenz (4) discuss conditions under which causality could be inferred from quantile regression. While causality is undoubtedly interesting, it was not the point of our paper. Our goal was to reconcile conflicting associations—not causal relationships—between happiness and income.

"Do larger incomes make people happier?" reads like a question regarding the causal effect of income on happiness, and KKM provide the answer that "the suffering of the unhappy group diminishes as income increases up to 100K but very little beyond that." For this to be true, the corresponding quantile treatment effects need to be *causally identified* (Fig. 1, box 1), and *rank invariance/similarity* (Fig. 1, box 2) needs to hold.

Causal inference problems

- » Researchers don't even realize or acknowledge that their research question requires causal reasoning
- » **Researchers have no idea how to connect their causal research questions to data and no idea which assumptions are involved**
- » Researchers draw causal inferences that rest on heroic assumptions

„Our study was only observational, and so no causal conclusions can be drawn. Future longitudinal studies are needed to determine whether...“

Literature on (observational) longitudinal data modeling

[1] While the omitted variable problem implies that we cannot make strong causal statements based on correlational data, it does not prohibit the use of the concept of *Granger causality* (Granger, 1969). However, many researchers using cross-lagged regression refrain from using the term *causal*, and use terms like *reciprocal relationship* (Erickson, Wolfe, King, King, & Sharkansky, 2001; Lindwall, Larsman, & Hagger, 2011), *role* (Ribeiroet al., 2011), *cross-domain effects* (Burt, Obradović, Long, & Masten, 2008), *exposure* (Cole et al., 2006), *impact* (Gault-Sherman, 2012), or *influence* (Green, Furrer, & McAllister, 2011), instead. It may be argued however, that these alternative terms also imply a causal mechanism, and even more so, that an interest in causality is actually the driving force behind these studies. Therefore, we decided to use the terms *causal* and *causality* in the current article, although we acknowledging that strong causal statements can only be based on experimental designs, and we should confine ourselves to the concept of Granger causality.

Hamaker et al. (2015). A Critique of the Cross-Lagged Panel Model.

Literature on (observational) longitudinal data modeling

[1] While the omitted variable problem implies that we cannot make strong causal statements based on correlational data, it does not prohibit the use of the concept of *Granger causality* (Granger, 1969). However, many researchers using cross-lagged regression refrain from using the term *causal*, and use terms like *reciprocal relationship* (Erickson, Wolfe, King, King, & Sharkansky, 2001; Lindwall, Larsman, & Hagger, 2011), *role* (Ribeiroet al., 2011), *cross-domain effects* (Burt, Obradović, Long, & Masten, 2008), *exposure* (Cole et al., 2006), *impact* (Gault-Sherman, 2012), or *influence* (Green, Furrer, & McAllister, 2011), instead. It may be argued however, that these alternative terms also imply a causal mechanism, and even more so, that an interest in causality is actually the driving force behind these studies. Therefore, we decided to use the terms *causal* and *causality* in the current article, although we acknowledging that strong causal statements can only be based on experimental designs, and we should confine ourselves to the concept of Granger causality.

Hamaker et al. (2015). A Critique of the Cross-Lagged Panel Model.

Causal inference problems

- Researchers don‘t even realize or acknowledge that their research question requires causal reasoning
- Researchers have no idea how to connect their causal research questions to data and no idea which assumptions are involved
- Researchers draw causal inferences that rest on heroic assumptions
  - Observational & experimental edition

Causal inference problems

- Researchers don't even realize or acknowledge that their research question requires causal reasoning
- Researchers have no idea how to connect their causal research questions to data and no idea which assumptions are involved
- Researchers draw causal inferences that rest on heroic assumptions
  - Observational & experimental edition

How can we do better?

A different type of causal ladder

Cutting edge, people working on the development of new methods

Translational methodologists

Substantive researcher who has never been properly introduced to causal inference

Communities and incentive structures

Division of cognitive labor, specialization

Alex Ware on Unsplash

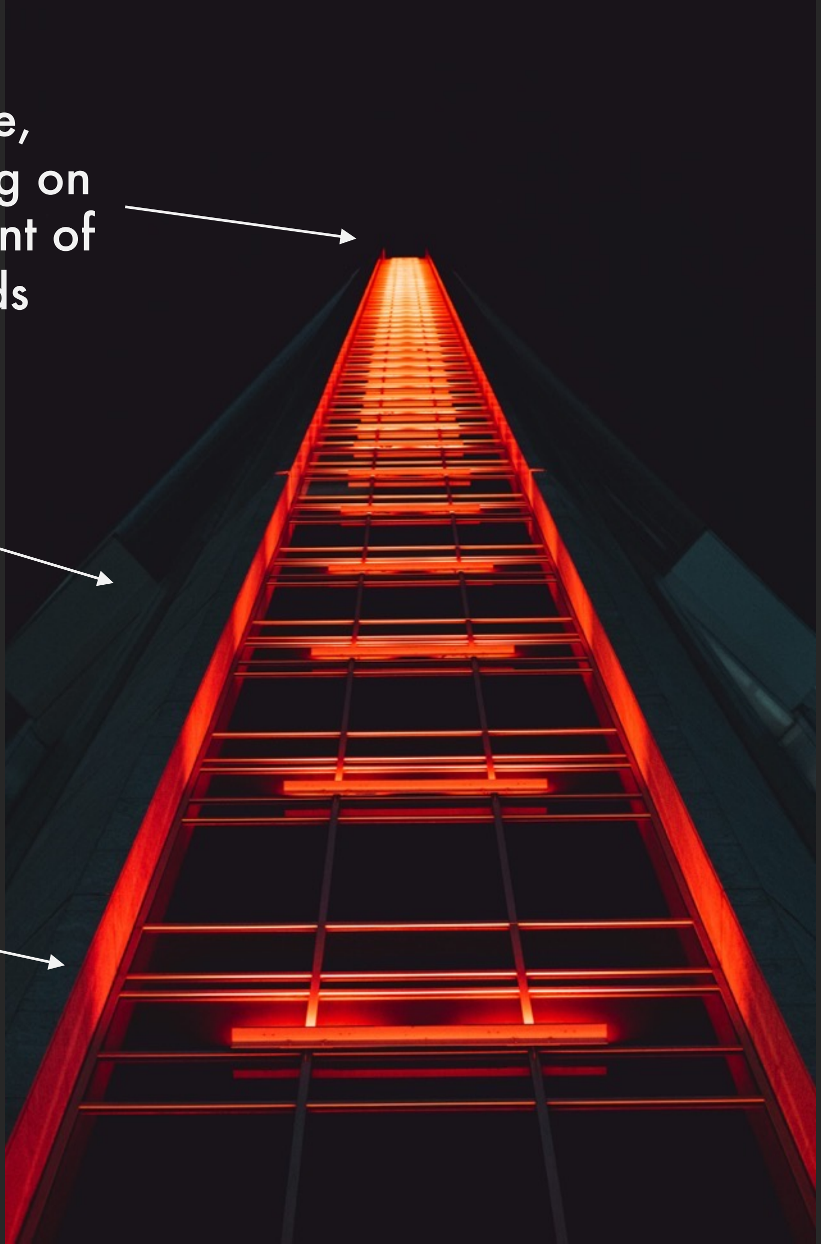
Slides: juliarohrer.com/resources

A different type of causal ladder

Cutting edge,
people working on
the development of
new methods

Translational
methodologists

Substantive researcher
who has never been
properly introduced to
causal inference



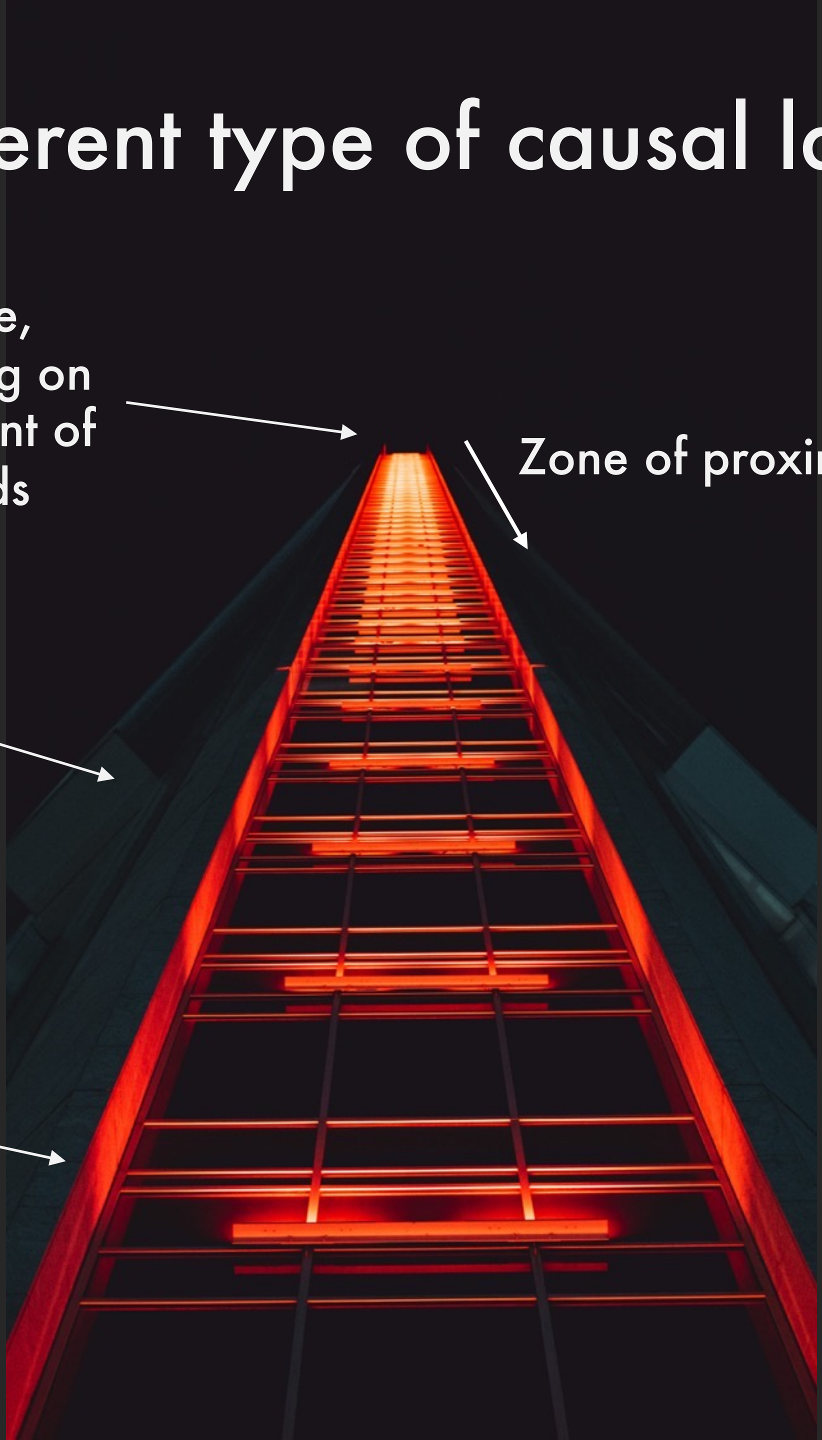
Communities and incentive
structures

Division of cognitive labor,
specialization

Alex Ware on Unsplash

Slides: juliarohrer.com/resources

A different type of causal ladder



Cutting edge, people working on the development of new methods

Translational methodologists

Substantive researcher who has never been properly introduced to causal inference

Zone of proximal outreach

- Contribute to the quality of the scientific output of the whole community
- Ensure that your methods are actually used and make a difference in practice

Alex Ware on Unsplash

A different type of causal ladder

Cutting edge, people working on the development of new methods

Translational methodologists

Substantive researcher who has never been properly introduced to causal inference

Zone of proximal outreach

- Contribute to the quality of the scientific output of the whole community
- Ensure that your methods are actually used and make a difference in practice

Zone of proximal development

- Contribute to the quality of the scientific output of the whole community
- Ensure that your work isn‘t horribly wrong

Alex Ware on Unsplash

Some ideas for how to help others improve their causal inferences

» Provide resources

» Provide criticism

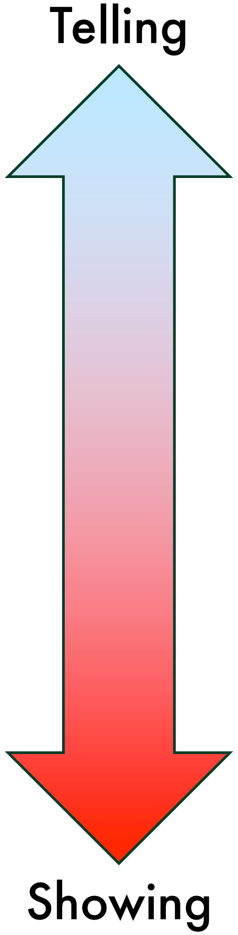
» Structural change (?)

Provide resources

» Tutorials & explainer articles

» Sneak tutorials

» Template articles



Tutorials & explainers – Some recommendations

» Tailoring it to a specific (sub-)field may greatly increase actual uptake

Tutorials & explainers – Some recommendations

» Tailoring it to a specific (sub-)field may greatly increase actual uptake

Equivalence testing for psychological research: A tutorial 1487 2018
D Lakens, AM Scheel, PM Isager
Advances in Methods and Practices in Psychological Science 1 (2), 259-269
+ Paperpile

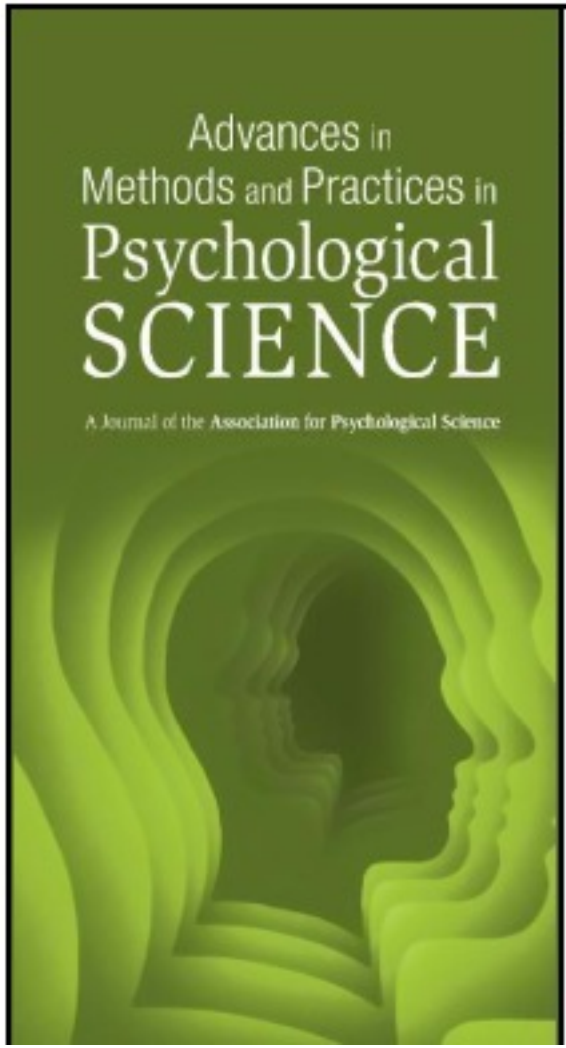
» Forget about novelty

Tutorials & explainers – Some recommendations

- Tailoring it to a specific (sub-)field may greatly increase actual uptake

Equivalence testing for psychological research: A tutorial 1487 2018
D Lakens, AM Scheel, PM Isager
Advances in Methods and Practices in Psychological Science 1 (2), 259-269
Paperpile

- Forget about novelty
- Consideration regarding a suitable outlet
  - More senior researchers may be more likely to pay attention (and cite) if it's an actual journal article in an established outlet



SUBMIT YOUR WORK

Advances in Methods and Practices in Psychological Science
Submission Site Now Open

Advances in Methods and Practices in Psychological Science (*AMPPS*) brings methodological advances to psychological scientists at-large. *AMPPS* seeks submissions that are accessible to and representative of the broad research interests of the field, including:

- Articles that communicate advances in methods, practices, and meta-science
- Empirical research that exemplifies scientific best practices
- Tutorials, commentaries, and simulation studies for new techniques and research tools
- Papers that bring advances from a specialized subfield to a broader audience
- Registered replication reports

Submit your work for consideration today for your chance to help APS lead the charge in advancing research methods in the field of psychological science.

Tutorials & explainers – Some recommendations

» Accessibility is key
  » But also hard (theory of mind task)
  » It may make perfect sense to team up with a substantive researcher, even if they don't quite understand (yet) what you are writing about
  » Some writing advice: „Writing about technical topics in an accessible manner“
  » My offer to you: if you're writing something for a non-technical psych audience, send it my way and I will take a look

Sneak tutorials

- A substantive article that is actually a tutorial in disguise
- Strengths
  - More likely to be actually read by substantive researchers (who work on that particular topic)
  - Potential to become a modern classic
    - People actually enjoy reading articles that make them feel like they learnt something important

RESEARCH ARTICLE | JULY 29 2024

The Effects of Satisfaction With Different Domains of Life on General Life Satisfaction Vary Between Individuals (but We Cannot Tell You Why)

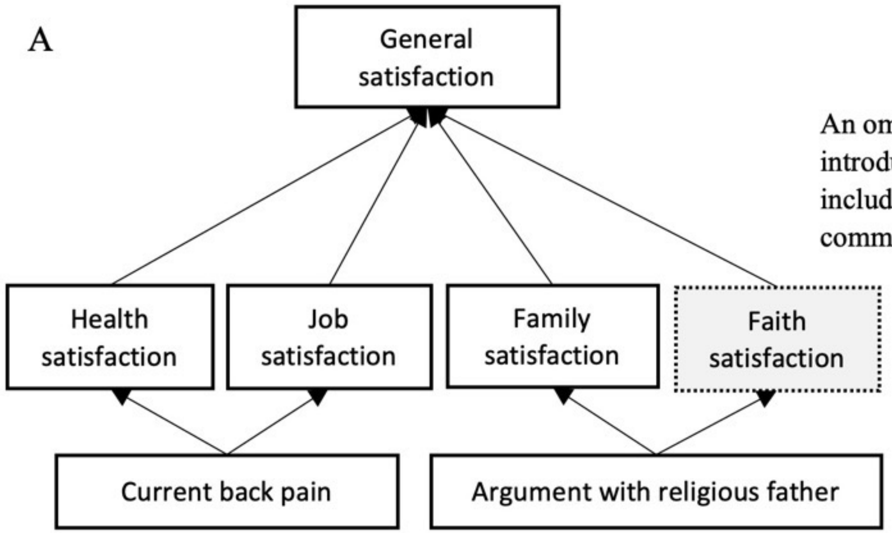
Collections: Section: Personality Psychology

Julia Rohrer, Ingo S. Seifert, Ruben C. Arslan, Jessie Sun, Stefan C. Schmukle

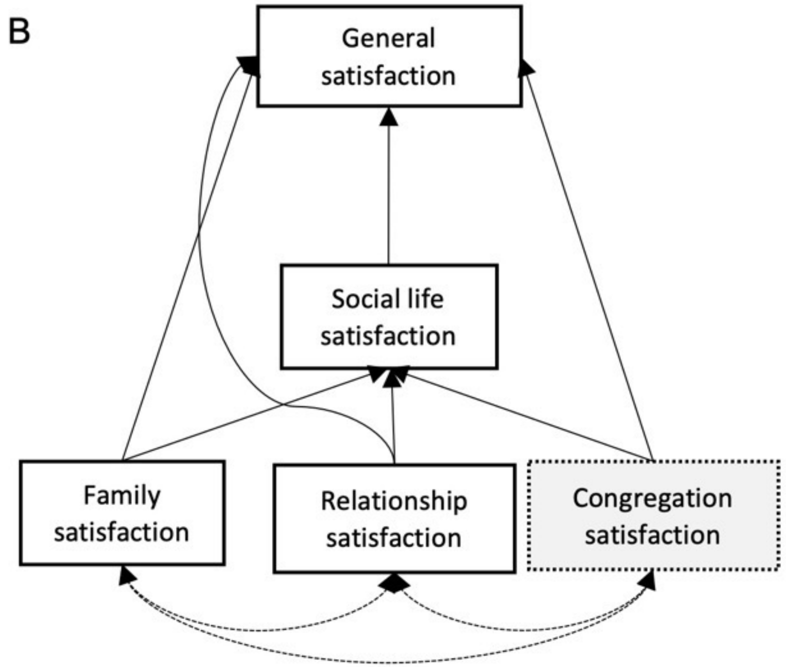
Editor: Yanna Weisberg

Corresponding author: julia.rohrer@uni-leipzig.de

Collabra: Psychology (2024) 10 (1): 121238.
https://doi.org/10.1525/collabra.121238 **Article history**



An omitted life domain (faith) can introduce confounding bias between an included domain (family) if they share common causes.



Conditioning on a higher-level domain introduces non-causal associations between lower-level domains. If one o those lower-level domains has been omitted from the analysis, this will bia the estimated direct effects of the lowe level domains.

https://online.ucpress.edu/collabra/article/10/1/121238/2884/The-Effects-of-Satisfaction-With-Different-Domains

Template articles

- A substantive article that handles the analyses and their justification really really well – so that it can serve as a template for other substantive researchers
- This is urgently needed for many causal inference issues
- Potential impact is huge
  - Researchers use existing publications (preferably in prestigious outlets…) as templates to structure their own work
- For this, you absolutely need to team up with a domain expert

JOURNAL ARTICLE

Assessing age trajectories (of subjective well-being): clarifying estimands, identification assumptions, and estimation strategies

Fabian Kratz, Josef Brüderl

European Sociological Review, jcaf038, https://doi.org/10.1093/esr/jcaf038

Published: 23 September 2025 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

Assessments of age trajectories are crucial for various strands of sociological research. This study provides guidance on how to estimate age trajectories by discussing a research design to answer the question, 'How does aging affect subjective well-being?' We define the estimand, discuss key identification and estimation assumptions, and propose—informed by this discussion—a research design. By contrasting our design with those of previous studies, we conclude that many prior investigations were affected by three biases related to identification and one bias concerning estimation. Using extensive panel data from the German Socio-Economic Panel Study (SOEP), we demonstrate that these four biases contributed to mixed empirical evidence, producing distortions that led to even qualitatively different conclusions. Finally, we discuss implications of our study for life course research and for enhancing the credibility of social science more broadly.

Issue Section: Original Article

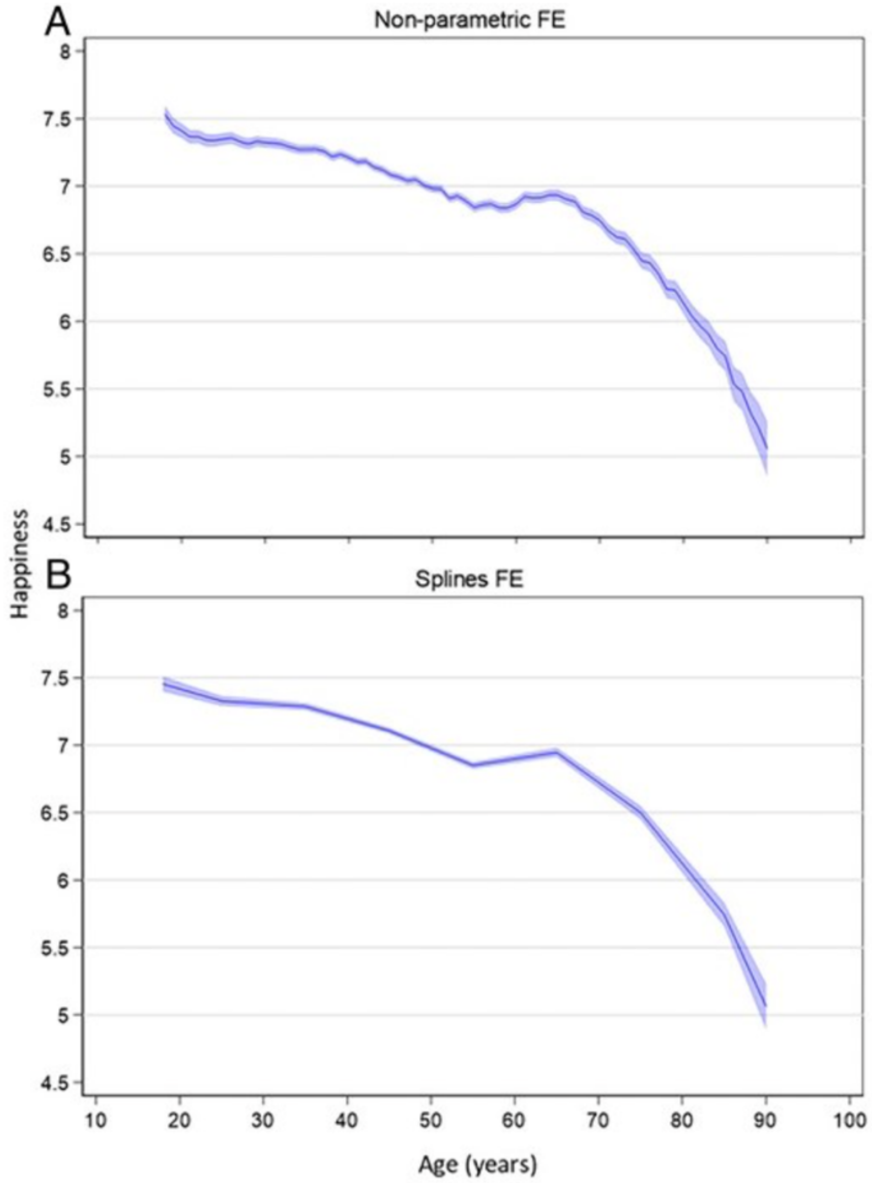


Figure 5 Predicted age–happiness trajectories (including 95 per cent-CIs) resulting from our best-practice design. *Notes:* (A) Predicted happiness values resulting from an FE model. Supplementary Table S6 (model 3) shows estimation details. (B) Predicted happiness values resulting from a spline FE model. Supplementary Table S7 shows estimation details. *Data source:* SOEP v34, own computations

Some ideas on what you can do

- » Provide resources
- » Provide criticism
- » Structural change (?)

Criticism: Commentaries

- For example, pointing out a common causal inference issue affecting a substantive high profile publication
  - If you see something, say something
- Productive framing: Point out issue in specific paper, then broaden – issue actually affects large swaths of the literature
  - Less of an attack on the authors
  - Ensures appropriate generalization on the side of the reader

Criticism: Commentaries

- Disclaimer: Some people have made frustrating experiences trying to publish commentaries
- Staged approach
  - Try original outlet (some journals are quite receptive, e.g., *Psychological Science*)

Criticism: Commentaries

- Disclaimer: Some people have made frustrating experiences trying to publish commentaries
- Staged approach
  - Try original outlet (some journals are quite receptive, e.g., *Psychological Science*)
  - Consider alternative peer-reviewed outlet (e.g., *Meta-Psychology* explicitly publishes commentaries on articles published in other journals)
  - Cut losses if necessary
    - Publish as blog post, preprint, on PubPeer

Criticism: Peer-review

Collective Intelligence
Volume 0: 1–18

DOI: 10.1177/26339137221131816
journals.sagepub.com/home/col
SAGE

Interdisciplinarity can aid the spread of better methods between scientific communities

Paul E Smaldino
University of California, Merced, CA, USA; Santa Fe Institute, Santa Fe, NM, USA

Cailin O'Connor
University of California, Irvine, CA, USA

Abstract
Why do bad methods persist in some academic disciplines, even when they have been widely rejected in others? What factors allow good methodological advances to spread across disciplines? In this paper, we investigate some key features determining the success and failure of methodological spread between the sciences. We introduce a formal model that considers factors like methodological competence and reviewer bias toward one's own methods. We show how these self-preferential biases can protect poor methodology within scientific communities, and lack of reviewer competence can contribute to failures to adopt better methods. We then use a second model to argue that input from outside disciplines can help break down barriers to methodological improvement. In doing so, we illustrate an underappreciated benefit of interdisciplinarity.

https://journals.sagepub.com/doi/full/10.1177/26339137221131816

Criticism: Peer-review

- You all have expertise that is tremendously valuable to editors at more applied journals
- Major problem: scalability
  - Focus on journals that are actually read and influence practices in the field
  - *Explicitly* just review individual parts of the methodology

THE 100% CI

ALWAYS CONFIDENT. SOMETIMES CREDIBLE.

Reviewer notes: In a randomized experiment, the pre-post differences are not effect estimates

January 22, 2025 · Julia Rohrer · 4 Comments

Reviewer notes are a new short format with brief explanations of basic ideas that might come in handy during (for...

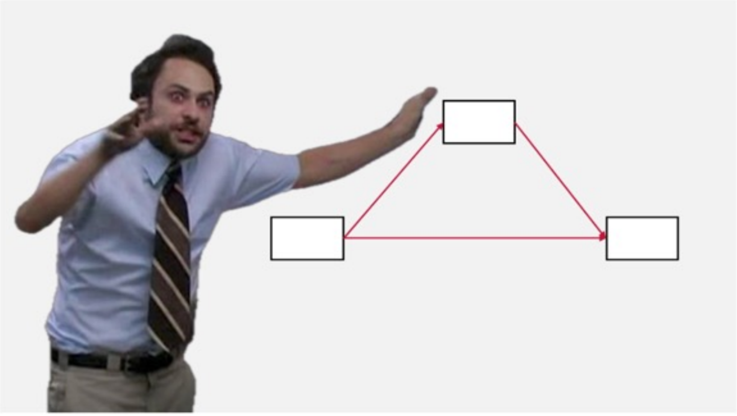
Continue Reading →

Reviewer notes: Avoid any ambiguity about analysis aims

February 17, 2025 · Julia Rohrer

For any central statistical analysis that you report in your manuscript, it should be absolutely clear for readers why the...

Continue Reading →



Reviewer notes: That's a very nice mediation analysis you have there. It would be a shame if something happened to it.

March 20, 2025 · Julia Rohrer · 13 Comments

Mediation analysis has gotten a lot of flak, including classic titles such as "Yes, but what's the mechanism? (Don't expect...

Continue Reading →

https://www.the100.ci/

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

Institute of Public Health

EN

- Research
- Academic Programs
- Back
- About the Institute
- Careers
- Network
- Go to Charité-Portal

Peerspectives

Peer Review Training Initiative for the Biomedical Sciences

Curious about the world of scientific peer review in the biomedical sciences? Want to gain mentor-guided, hands-on reviewer experience while contributing to scientific quality control in a tangible way? Interested in gaining insights from experienced experts about what peer reviewers and journal editors from top journals are looking for? Eager to learn how you can improve your own writing and reporting practice?

Search this Website

Home › Academic Programs › PhD in Health Data Sciences › Peerspectives

About the Program

Peerperspectives participants gain valuable perspectives on these and other topics related to scientific peer review and beyond!

By creating Peerspectives, our aim was to help early career researchers better understand the intricacies of the

peerspectives

https://iph.charite.de/en/academic_programs/phd_in_health_data_sciences/peerspectives/

ERЯOR: A Bug Bounty Program for Science

ERЯOR is a comprehensive program to systematically detect and report errors in scientific publications, modelled after bug bounty programs in the technology industry. Investigators are paid for discovering errors in the scientific literature: The more severe the error, the larger the payout. In ERЯOR, we leverage, survey, document, and increase accessibility to error detection tools. Our goal is to foster a culture that is open to the possibility of error in science to embrace a new discourse norm of constructive criticism.

4 / 11	151	4	5'550
			/ 250'000 CHF
Reviews Completed / Pending	Invited Papers	Papers with Errors	Total Payouts

Reviewers needed!

Even more so if you have SAS knowledge...

More structural change

- Changing training from scratch
  - e.g., research methods training with causal graphs
  - Sneak causality into other courses

More structural change

- Changing training from scratch
  - e.g., research methods training with causal graphs
  - Sneak causality into other courses
- Changing publication requirements
  - Structured abstracts with transparent estimands and identification assumptions (Who would win, 100 duck-sized strategic ambiguities vs. 1 horse-sized structured abstract?)

More structural change

- Changing training from scratch
  - e.g., research methods training with causal graphs
  - Sneak causality into other courses
- Changing publication requirements
  - Structured abstracts with transparent estimands and identification assumptions (Who would win, 100 duck-sized strategic ambiguities vs. 1 horse-sized structured abstract?)
- Changing how researchers are assessed (PhD criteria, hiring...)
  - Rewarding quality over quantity (Towards responsible research assessment: How to reward research quality)

Thank you for your attention!

Slides →

Julia M. Rohrer
Leipzig University
www.juliarohrer.com
@dingdingpeng.the100.ci
www.the100.ci