This mess we're in?

Or how simulation and prediction will advance (demographic) research

gert stulp — gertstulp.com — g.stulp@rug.nl

How Well Are We Doing?

variables explain little

Fewer births through education and flexwork?

"total effect on fertility … rather small

variables explain little

Fewer births through education and flexwork?

"total effect on fertility … rather small

surprising patterns

variables explain little

Fewer births through education and flexwork?

“total effect on fertility … rather small

surprising patterns

incomparable results


PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Review

Wealth, fertility and adaptive behaviour in industrial populations

Gert Stulp and Louise Barrett

1. Introduction

THE ROYAL SOCIETY

…ieties, Part I

…cieties, Part II

variables explain little

Fewer births through education and flexwork?

“total effect on fertility … rather small

surprising patterns

incomparable results

non-replicable results

My Upbringing in Science

2008

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

Daryl J. Bem

Experiencing Physical Warmth Promotes Interpersonal Warmth

Lawrence E. Williams and John A. Bargh

Hot drinks encourage warmer feelings

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

General Article

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

Joseph P. Simmons, Leif D

Journal of Experimental Psychology: General 2014, Vol. 143, No. 2, 534–547

P-Curve: A Key to the File-Drawer

Uri Simonsohn, University of Pennsylvania
Leif D. Nelson, University of California, Berkeley
Joseph P. Simmons, University of Pennsylvania

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

The natural selection of bad science

Canadian Psychology / Psychologie canadienne

2020, Vol. 61, No. 4, 281–288

Psychological Measurement and the Replication Crisis: Four Sacred Cows

Scott
Emory University

Perspectives on Psychological Science 2021, Vol. 16(4) 779–788

The Theory Crisis in Psychology: How to Move Forward

Markus I. Eronen and Laura F. Bringmann

Department of Theoretical Philosophy, and Department of Psychometrics and Statistics, University of Groningen

Science

Pacific Standard

SCIENTIFIC DATA

OPEN

Assessing data availability and research reproducibility in hydrology and water resources

My Upbringing in Science

2008

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

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

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. The term is purely descriptive; it neither implies that such phenomena are paranormal nor connotes anything about their underlying mechanisms. Alleged psi phenomena include *telepathy*, the apparent transfer of information from one person to another without the mediation of any known channel of sensory communication; *clairvoyance* (sometimes called *remote viewing*), the apparent perception of objects or events that do not provide a stimulus to the known senses; *psychokinesis*, the apparent influence of thoughts or intentions on physical or biological processes; and *precognition* (conscious cognitive awareness) or *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 nine experiments designed to test for such retroactive influence by "time-reversing" several well-established psychological effects, so that the individual's responses are obtained before the putatively causal stimulus events occur.

Psi is a controversial subject, and most academic psychologists do not believe that psi phenomena are likely to exist. A survey of 1,100 college professors in the United States found that psychologists were much more skeptical about the existence of psi than were their colleagues in the natural sciences, the other social sciences, or the humanities (Wagner & Monnet, 1979). In fact, 34% of the psychologists in the sample declared psi to be impossible, a view expressed by only 2% of all other respondents. Although our colleagues in other disciplines would probably agree with the oft-quoted dictum that "extraordinary claims require extraordinary evidence," we psychologists are more likely to be familiar with the methodological and statistical requirements for sustaining such claims and aware of previous claims that failed either to meet those requirements or to survive the test of successful replication. Several other reasons for our greater skepticism are discussed by Bem and Honorton (1994, pp. 4–5).

There are two major challenges for psi researchers, one empirical and one theoretical. The major empirical challenge, of course, is to provide well-controlled demonstrations of psi that can be replicated by independent investigators. That is the major goal in the research program reported in this article. Accordingly, the

This article was published Online First January 31, 2011.

I am grateful to the students who served as head research assistants and laboratory coordinators for their enthusiasm and dedication to this controversial enterprise: Ben Edelman, Rebecca Epstein, Dan Fishman, Jamison Hahn, Eric Hoffman, Kelly Lin, Brianne Mintern, Brittany Terner, and Jade Wu. I am also indebted to the 30 other students who served as friendly and reliable experimenters over the course of this research program. Dean Radin, senior scientist at the Institute of Noetic Sciences (IONS), and David Sherman, professor of psychology at the University of California, Santa Barbara, provided valuable guidance in the preparation of this article.

Correspondence concerning this article should be addressed to Daryl J. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853. E-mail: d.bem@cornell.edu

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

The natural selection of bad science

Canadian Psychology / Psychologie canadienne

Experiencing Physical Warmth Promotes Interpersonal Warmth

Lawrence E. Williams[1]* and John A. Bargh[2]

"Warmth" is the most powerful personality trait in social judgment, and attachment theorists have stressed the importance of warm physical contact with caregivers during infancy for healthy relationships in adulthood. Intriguingly, recent research in humans points to the involvement of the insula in the processing of both physical temperature and interpersonal warmth (trust) information. Accordingly, we hypothesized that experiences of physical warmth (or coldness) would increase feelings of interpersonal warmth (or coldness), without the person's awareness of this influence. In study 1, participants who briefly held a cup of hot (versus iced) coffee judged a target person as having a "warmer" personality (generous, caring); in study 2, participants holding a hot (versus cold) therapeutic pad were more likely to choose a gift for a friend instead of for themselves.

Hot drinks encourage warmer feelings

Curve: A

dard

TIFIC DATA

essing data availability and
arch reproducibility in
rology and water resources

My Upbringing in Science

2008

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

Experiencing Physical Warmth Promotes Interpersonal Warmth

Lawrence E. Williams and John A. Bargh

Hot drinks encourage warmer feelings

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

General Article

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

Joseph P. Simmons[1], Leif D…

Journal of Experimental Psychology: General

2014, Vol. 143, No. 2, 534–547

© 2013 American Psychological Association

P-Curve: A Key to the File-Drawer

Uri Simonsohn, University of Pennsylvania

Leif D. Nelson, University of California, Berkeley

Joseph P. Simmons, University of Pennsylvania

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

The natural selection of bad science

Canadian Psychology / Psychologie canadienne

2020, Vol. 61, No. 4, 281–288

Psychological Measurement and the Replication Crisis: Four Sacred Cows

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science

2021, Vol. 16(4) 779–788

The Theory Crisis in Psychology: How to Move Forward

Markus I. Eronen[1] and Laura F. Bringmann[2]

[1]Department of Theoretical Philosophy, and [2]Department of Psychometrics and Statistics, University of Groningen

Science

Pacific Standard

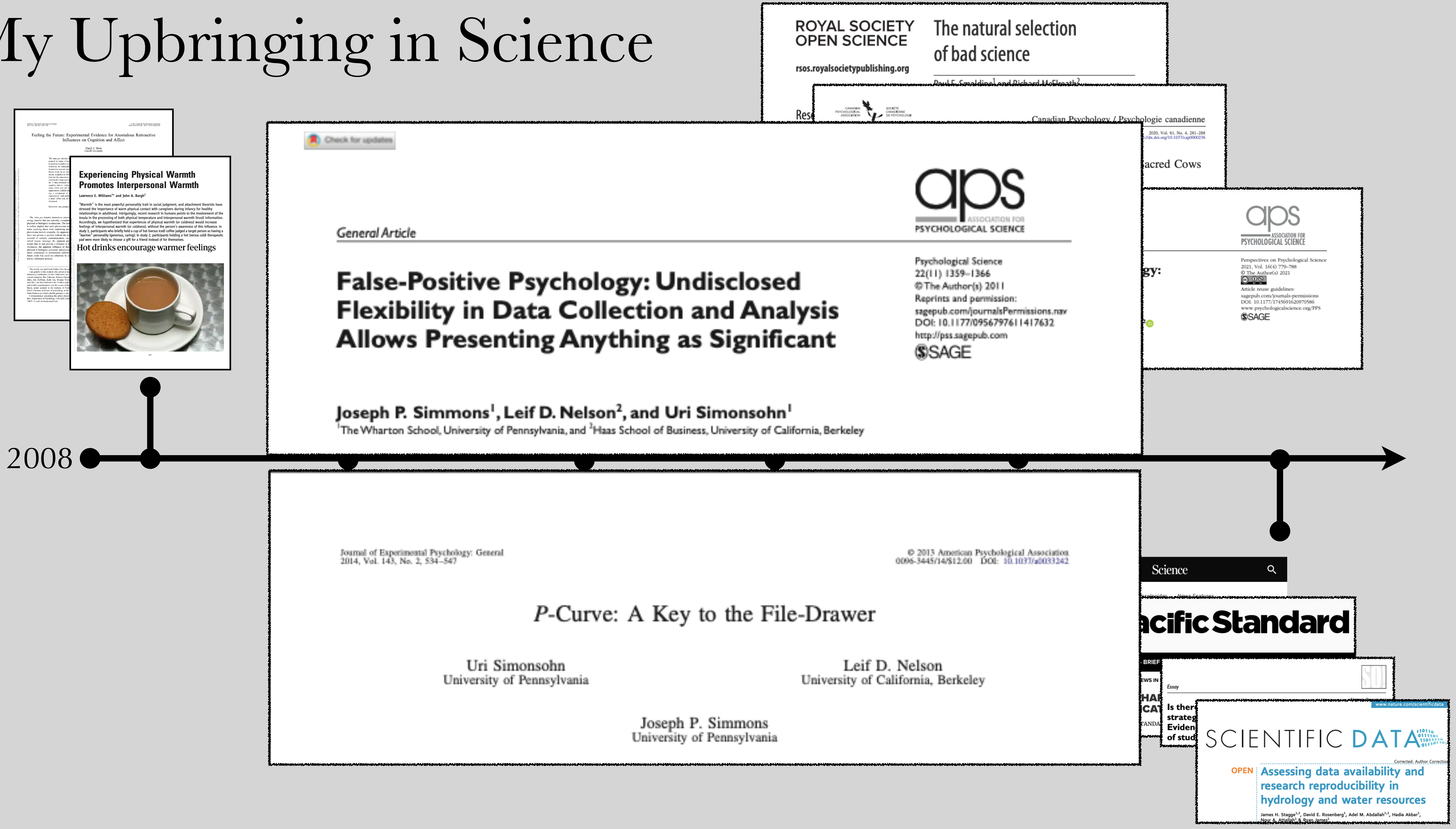
SCIENTIFIC DATA

OPEN

Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge, David E. Rosenberg, Adel M. Abdallah, Hadia Akbar

My Upbringing in Science



My Upbringing in Science

2008

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

Daryl J. Bem

Experiencing Physical Warmth Promotes Interpersonal Warmth

Lawrence E. Williams and John A. Bargh

"Warmth" is the most powerful personality trait in social judgment, and attachment theorists have stressed the importance of warm physical contact with caregivers during infancy for healthy relationships in adulthood. Intriguingly, recent research in humans points to the involvement of the insula in the processing of both physical temperature and interpersonal warmth (trust) information. Accordingly, we hypothesized that experiences of physical warmth (or coldness) would increase feelings of interpersonal warmth (or coldness), without the person's awareness of this influence. In study 1, participants who briefly held a cup of hot (versus iced) coffee judged a target person as having a "warmer" personality (generous, caring); in study 2, participants holding a hot (versus cold) therapeutic pad were more likely to choose a gift for a friend instead of for themselves.

Hot drinks encourage warmer feelings

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

General Article

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

Joseph P. Simmons[1], Leif D...

Psychological Science 22(11) 1359–1366 © The Author(s) 2011

Journal of Experimental Psychology: General 2014, Vol. 143, No. 2, 534–547 © 2013 American Psychological Association

P-Curve: A Key to the File-Drawer

Uri Simonsohn, University of Pennsylvania

Leif D. Nelson, University of California, Berkeley

Joseph P. Simmons, University of Pennsylvania

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

The natural selection of bad science

Paul E. Smaldino[1] and Richard McElreath[2]

Canadian Psychology / Psychologie canadienne

ISSN: 0708-5591 © 2020 Canadian Psychological Association

2020, Vol. 61, No. 4, 281–288 http://dx.doi.org/10.1037/cap0000236

Psychological Measurement and the Replication Crisis: Four Sacred Cows

Scott ... Emory University

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science 2021, Vol. 16(4) 779–788 © The Author(s) 2021

The Theory Crisis in Psychology: How to Move Forward

Markus I. Eronen[1] and Laura F. Bringmann[2]

[1]Department of Theoretical Philosophy, and [2]Department of Psychometrics and Statistics, University of Groningen

Science

About 40... replicati...

Pacific Standard

NEWS IN BRIEF

BIG PHAR... REPLICAT...

Essay

Is ther... strateg... Eviden... of stud...

SCIENTIFIC DATA

OPEN

Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge, David E. Rosenberg, Adel M. Abdallah, Hadia Akbar, Nour A. Attallah & Ryan James

My Upbringing in Science

2008

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

CrossMark click for updates

Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science.

The natural selection of bad science

Paul E. Smaldino[1] and Richard McElreath[2]

[1]Cognitive and Information Sciences, University of California, Merced, CA 95343, USA
[2]Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

PES, 0000-0002-7133-5620; RME, 0000-0002-0387-5377

Check for updates

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

The Theory Crisis in Psychology: How to Move Forward

Markus I. Eronen[1] and Laura F. Bringmann[2]
[1]Department of Theoretical Philosophy, and [2]Department of Psychometrics and Statistics, University of Groningen

Perspectives on Psychological Science
2021, Vol. 16(4) 779–788
© The Author(s) 2021
Article reuse guidelines: sagepub.com/journals-permissions
DOI: 10.1177/1745691620970586
www.psychologicalscience.org/PPS
SAGE

CANADIAN PSYCHOLOGICAL ASSOCIATION SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

Canadian Psychology / Psychologie canadienne

ISSN: 0708-5591
© 2020 Canadian Psychological Association

2020, Vol. 61, No. 4, 281–288
http://dx.doi.org/10.1037/cap0000236

Psychological Measurement and the Replication Crisis: Four Sacred Cows

Scott O. Lilienfeld
Emory University and University of Melbourne

Adele N. Strother
Emory University

cience

cific Standard

Essay

Is ther
strateg
Eviden
of stud

www.nature.com/scientificdata

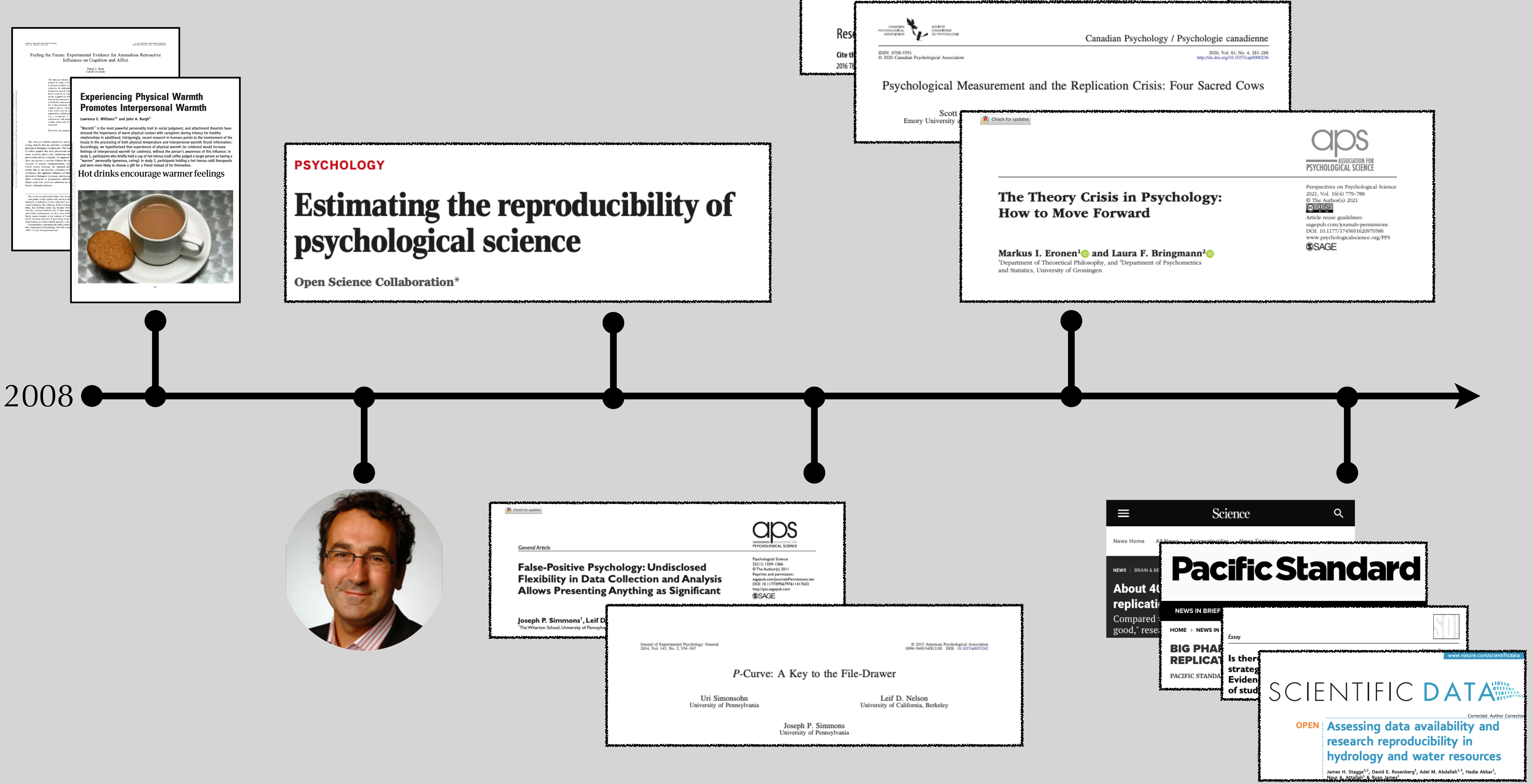
SCIENTIFIC DATA

Corrected: Author Correction

OPEN Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge[1,2], David E. Rosenberg[1], Adel M. Abdallah[1,3], Hadia Akbar[1], Nour A. Attallah[1] & Ryan James[1]

My Upbringing in Science



My Upbringing in Science

Science

News Home · All News · ScienceInsider · News Features

NEWS | BRAIN & BEHAVIOR

About 40% of economics experiments fail replication survey

Compared with psychology, the replication rate "is rather good," researchers say

ROYAL SOCIETY OPEN SCIENCE
rsos.royalsocietypublishing.org

The natural selection of bad science

Pacific Standard

NEWS IN BRIEF · ECONOMICS · EDUCATION · ENVIRONMENT · SOCIAL JUSTICE

HOME › NEWS IN BRIEF

BIG PHARMA REVEALS A BIOMEDICAL REPLICATION CRISIS

PACIFIC STANDARD STAFF · UPDATED: JUN 14, 2017 · ORIGINAL: MAY 12, 2016

2008

Essay

Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings

Strategic Organization
2017, Vol. 15(3) 423–436
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1476127017701076
journals.sagepub.com/home/soq
SAGE

www.nature.com/scientificdata

SCIENTIFIC DATA

Corrected: Author Correction

OPEN

Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge[1,2], David E. Rosenberg[1], Adel M. Abdallah[1,3], Hadia Akbar[1], Nour A. Attallah[1] & Ryan James[1]

Replication (crisis) in Ecology?

Reasons why not

Reasons why

Ecological Applications, 29(1), 2019, e01822

Open science, reproducibility, and transparency in ecology

Stephen M. Powers[1] and Stephanie E. Hampton

School of the Environment, Washington State University, Pullman, Washington 99164 USA

Citation: S. M. Powers, a
parency in ecology. Ecolog

PeerJ

Rate and success of study replication in ecology and evolution

Clint D. Kelly

Département des Sciences biologiques, Universit

Statistical Reports

Ecology, 97(10), 2016, pp. 2554–2561

Underappreciated problems of low replication in ecological field studies

Nathan P. Lemoine,[1] Ava Hoffman, Andrew J. Felton, Lauren Baur, Francis Chaves, Jesse Gray, Qiang Yu,[2] and Melinda D. Smith

Estimating the reproducibility of social learning research published between 1955 and 2018

Riana Minocher, Silke Atmaca, Claudia Bavero, Richard McElreath and Bret Beheim

“We thus outline clear measures to improve the reproducibility of research on the ecology and evolution of social behaviour.

Replication (crisis) in Demography?

Reasons why not

Reasons why

Replication (crisis) in Demography?

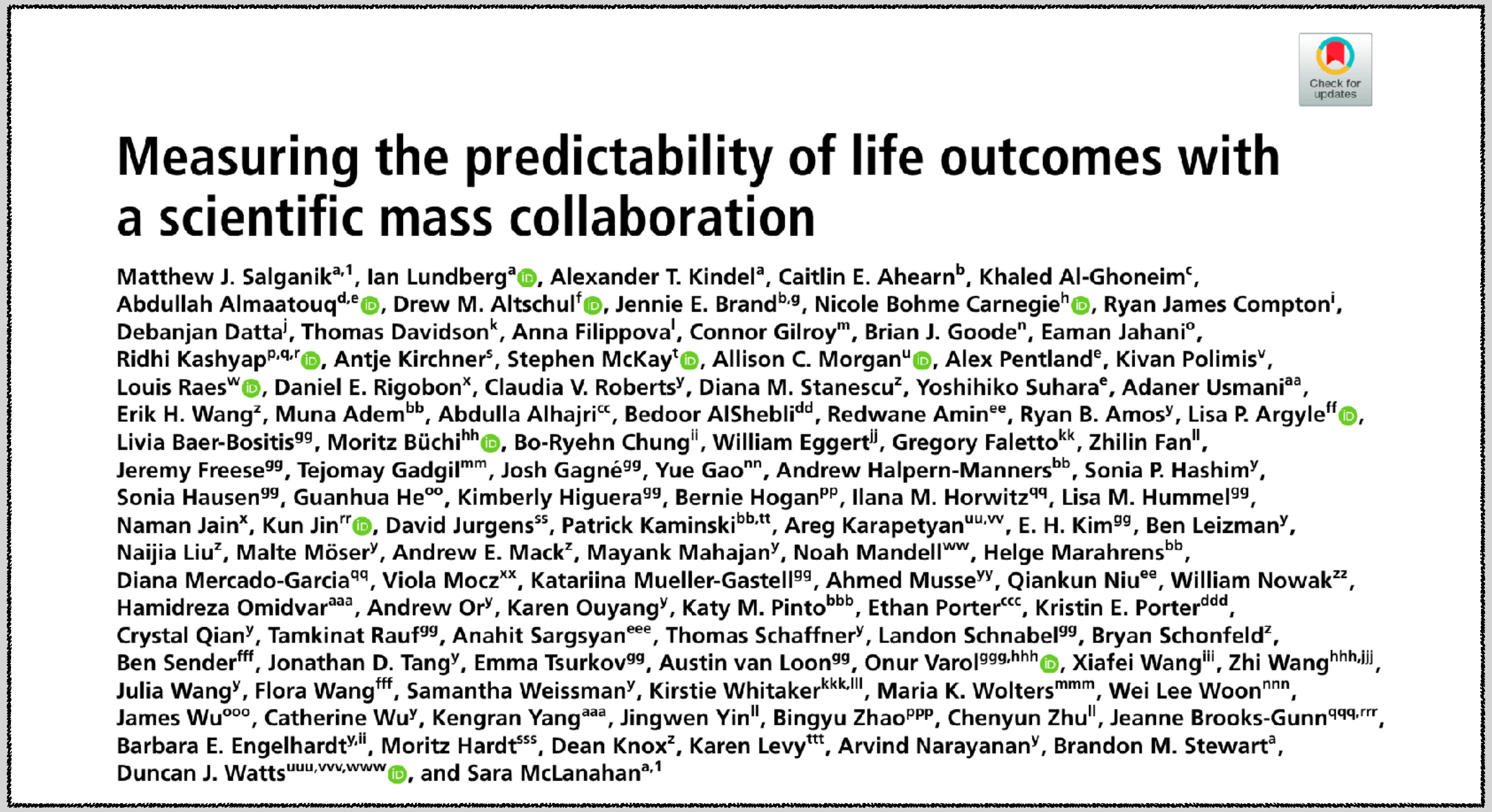
Reasons why not

- *Strong methods*
- *Strong focus on representative data*
- *Less measurement error*
- *Open data*
- *Large N*
- *Often descriptive*

Reasons why

- *Non-experimental*
- *Correlational, but little causal inference*
- *Large N, yet star gazing*
- *Controlling at will*
- *"Culture" as a get-out-of-jail-for-free card*

Predictability Crisis?

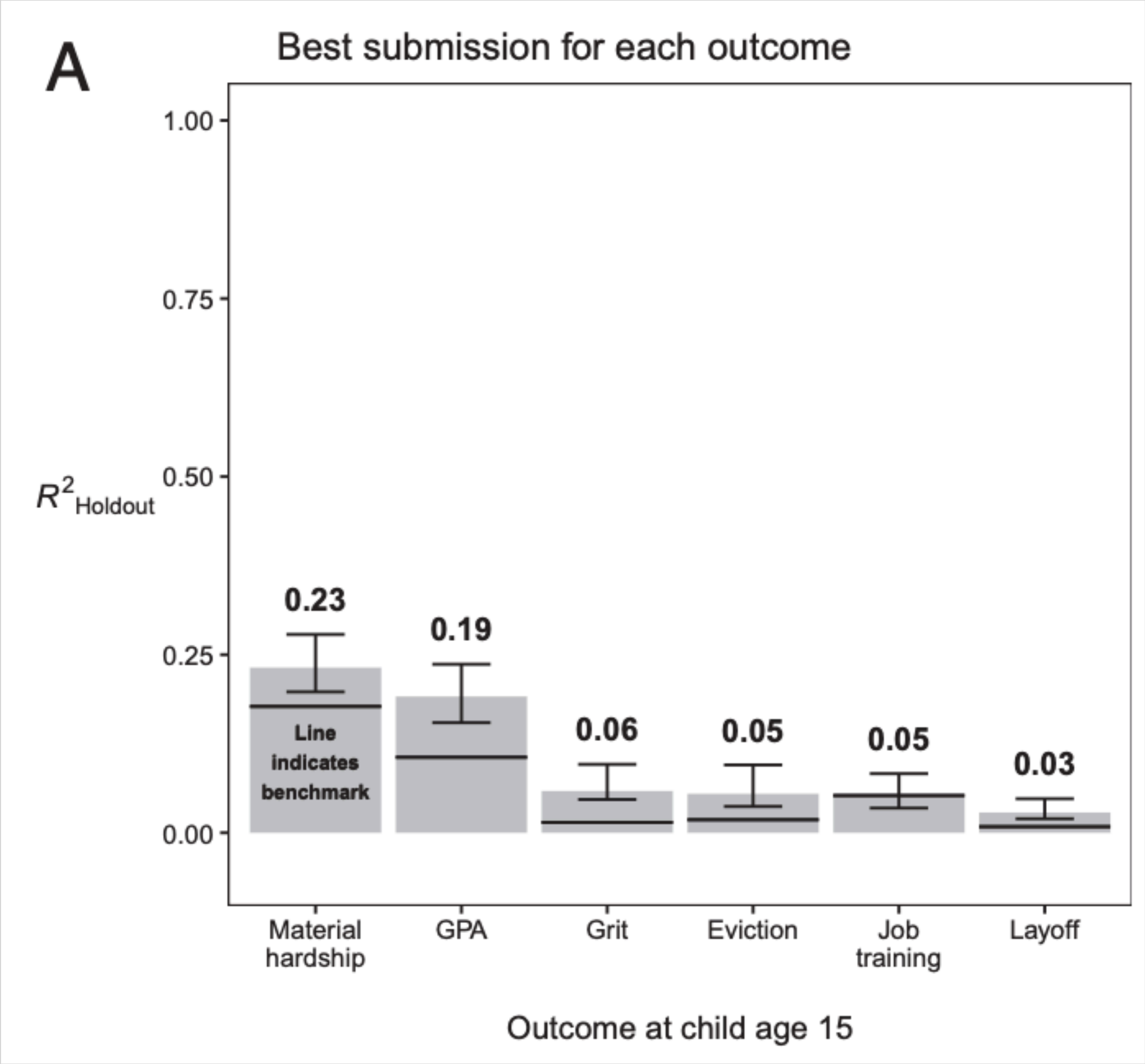


Check for updates

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik[a,1], Ian Lundberg[a], Alexander T. Kindel[a], Caitlin E. Ahearn[b], Khaled Al-Ghoneim[c], Abdullah Almaatouq[d,e], Drew M. Altschul[f], Jennie E. Brand[b,g], Nicole Bohme Carnegie[h], Ryan James Compton[i], Debanjan Datta[j], Thomas Davidson[k], Anna Filippova[l], Connor Gilroy[m], Brian J. Goode[n], Eaman Jahani[o], Ridhi Kashyap[p,q,r], Antje Kirchner[s], Stephen McKay[t], Allison C. Morgan[u], Alex Pentland[e], Kivan Polimis[v], Louis Raes[w], Daniel E. Rigobon[x], Claudia V. Roberts[y], Diana M. Stanescu[z], Yoshihiko Suhara[e], Adaner Usmani[aa], Erik H. Wang[z], Muna Adem[bb], Abdulla Alhajri[cc], Bedoor AlShebli[dd], Redwane Amin[ee], Ryan B. Amos[y], Lisa P. Argyle[ff], Livia Baer-Bositis[gg], Moritz Büchi[hh], Bo-Ryehn Chung[ii], William Eggert[jj], Gregory Faletto[kk], Zhilin Fan[ll], Jeremy Freese[gg], Tejomay Gadgil[mm], Josh Gagné[gg], Yue Gao[nn], Andrew Halpern-Manners[bb], Sonia P. Hashim[y], Sonia Hausen[gg], Guanhua He[oo], Kimberly Higuera[gg], Bernie Hogan[pp], Ilana M. Horwitz[qq], Lisa M. Hummel[gg], Naman Jain[x], Kun Jin[rr], David Jurgens[ss], Patrick Kaminski[bb,tt], Areg Karapetyan[uu,vv], E. H. Kim[gg], Ben Leizman[y], Naijia Liu[z], Malte Möser[y], Andrew E. Mack[z], Mayank Mahajan[y], Noah Mandell[ww], Helge Marahrens[bb], Diana Mercado-Garcia[qq], Viola Mocz[xx], Katariina Mueller-Gastell[gg], Ahmed Musse[yy], Qiankun Niu[ee], William Nowak[zz], Hamidreza Omidvar[aaa], Andrew Or[y], Karen Ouyang[y], Katy M. Pinto[bbb], Ethan Porter[ccc], Kristin E. Porter[ddd], Crystal Qian[y], Tamkinat Rauf[gg], Anahit Sargsyan[eee], Thomas Schaffner[y], Landon Schnabel[gg], Bryan Schonfeld[z], Ben Sender[fff], Jonathan D. Tang[y], Emma Tsurkov[gg], Austin van Loon[gg], Onur Varol[ggg,hhh], Xiafei Wang[iii], Zhi Wang[hhh,jjj], Julia Wang[y], Flora Wang[fff], Samantha Weissman[y], Kirstie Whitaker[kkk,lll], Maria K. Wolters[mmm], Wei Lee Woon[nnn], James Wu[ooo], Catherine Wu[y], Kengran Yang[aaa], Jingwen Yin[ll], Bingyu Zhao[ppp], Chenyun Zhu[ll], Jeanne Brooks-Gunn[qqq,rrr], Barbara E. Engelhardt[y,ii], Moritz Hardt[sss], Dean Knox[z], Karen Levy[ttt], Arvind Narayanan[y], Brandon M. Stewart[a], Duncan J. Watts[uuu,vvv,www], and Sara McLanahan[a,1]

data challenge:
predicting life outcomes
based on ~6000 variables
by 160 teams
both theory- & data-driven



Predictability Crisis?

> “Social scientists studying the life course must find a way to reconcile a widespread belief that understanding has been generated by these data—as demonstrated by more than 750 published journal articles using the Fragile Families data with the fact that the very same data could not yield accurate predictions of these important outcomes.

How Well Are We Doing?

The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

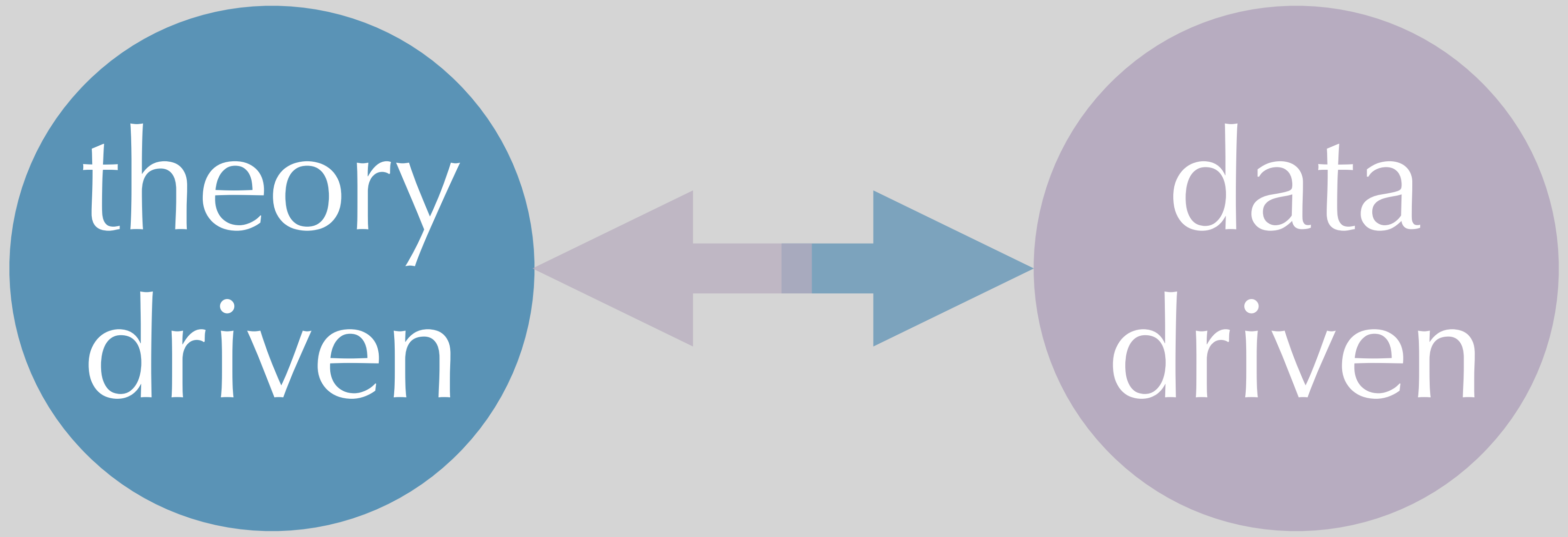
microsimulation can advance traditional statistical modelling

The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

out-of-sample predictive ability:

clear measure of effect size

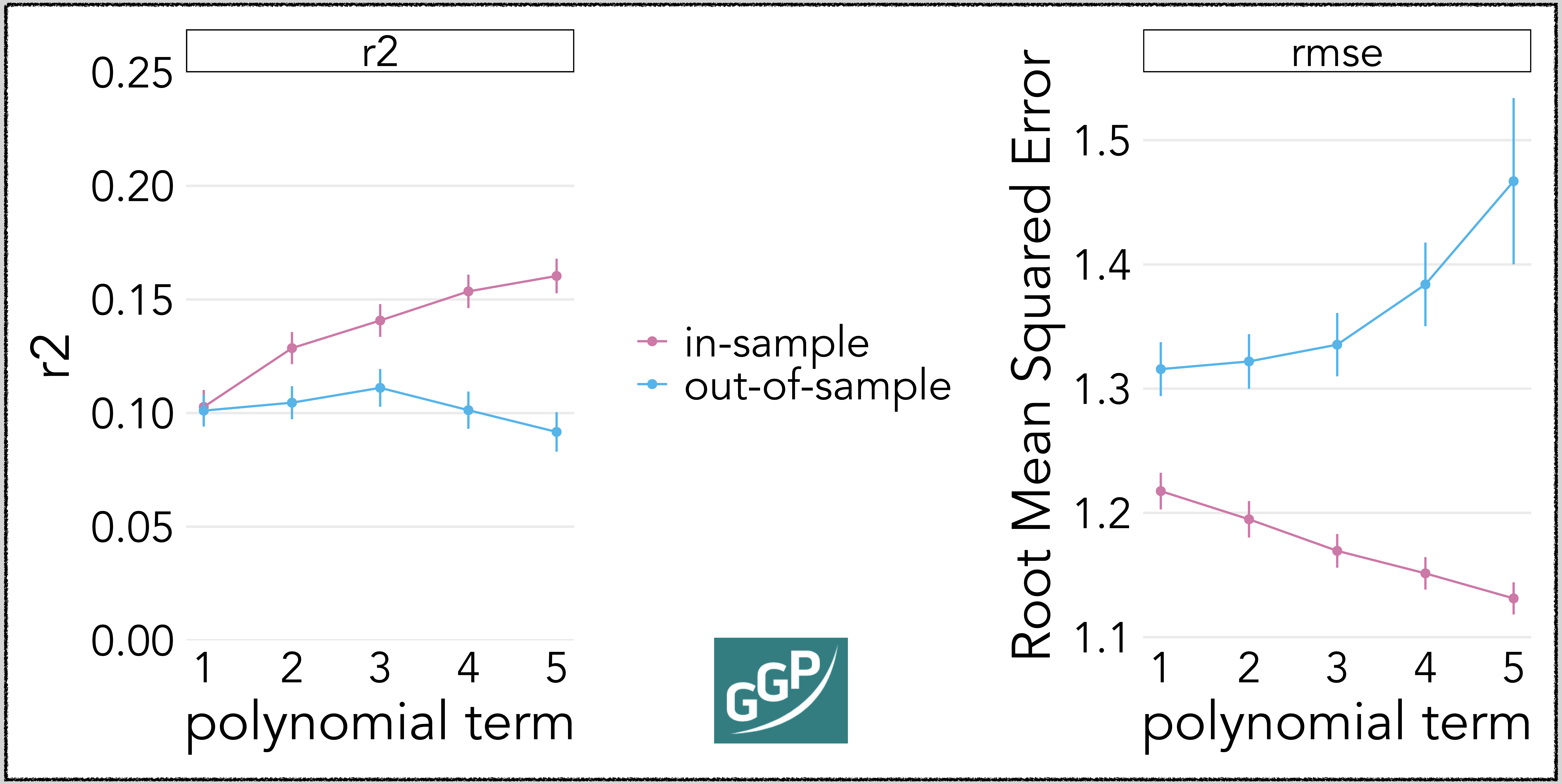


facilitates dialogue theory- and data-driven models

measure of distance theory and practice

Out-of-Sample Prediction

$n = 50$ *training data*
$n = 50$ *test data*
number of children ~ age

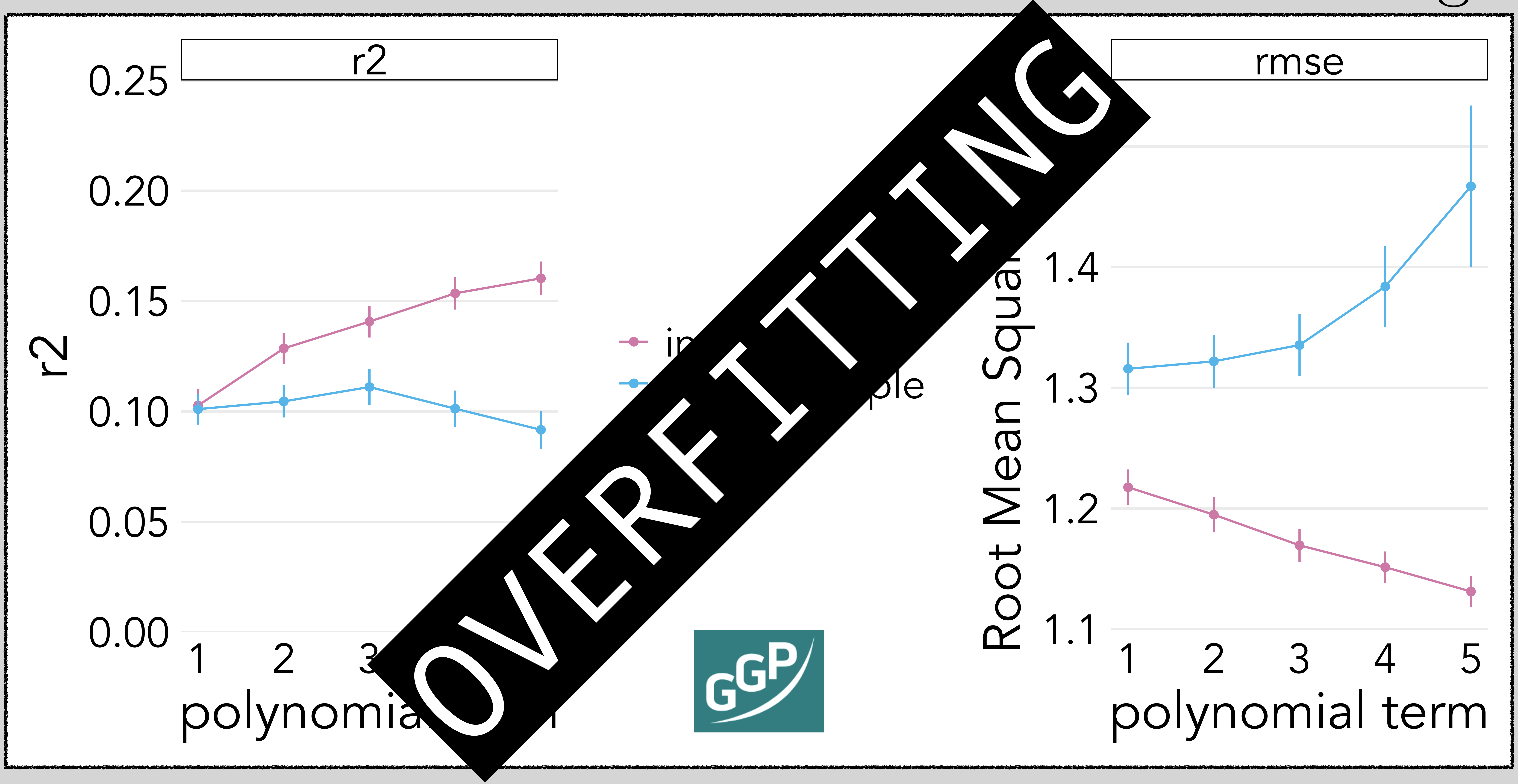


Out-of-Sample Prediction

$n = 50$ *training data*

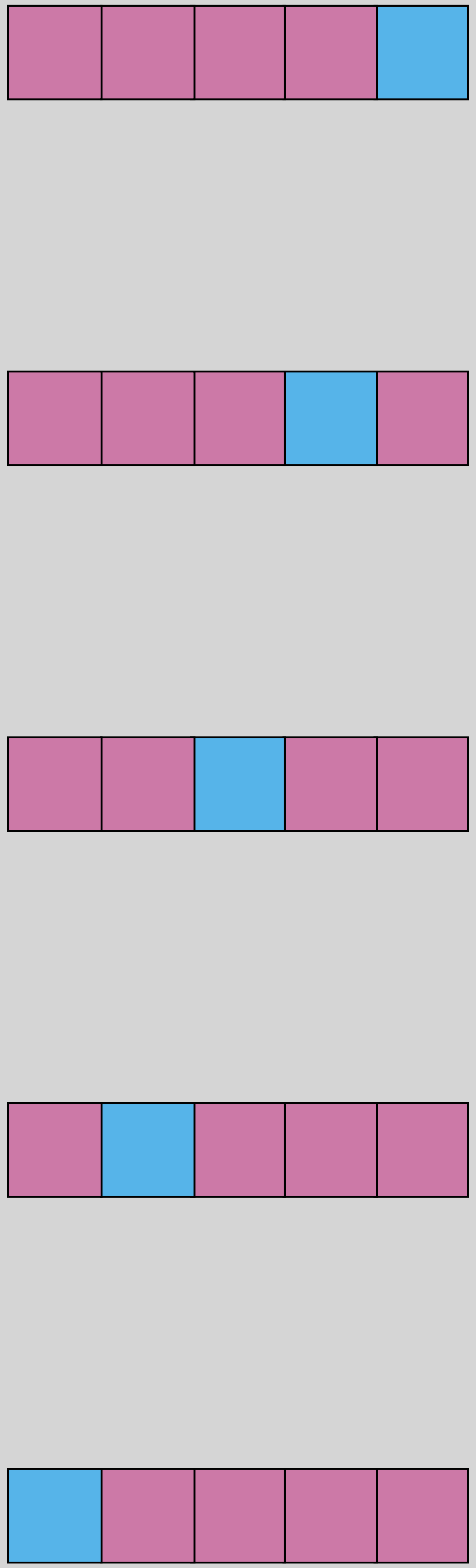
$n = 50$ *test data*

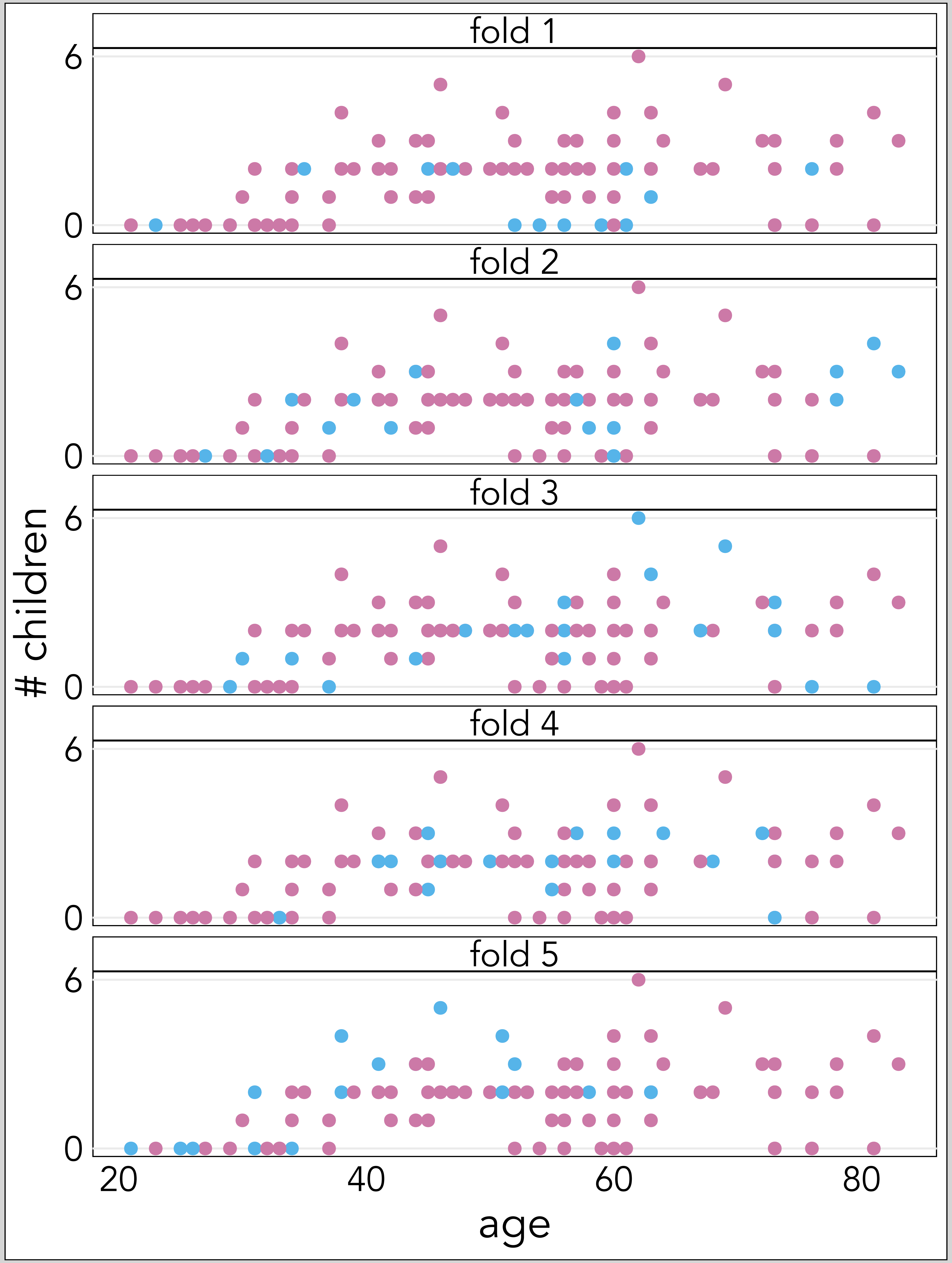
number of children ~ age



Cross Validation

fold	in-sample R^2	out-of-sample R^2
1	0.15	0.12
2	0.17	0.17
3	0.14	0.20
4	0.20	-0.18
5	0.27	-1.38
Average out-of-sample R^2		-0.07



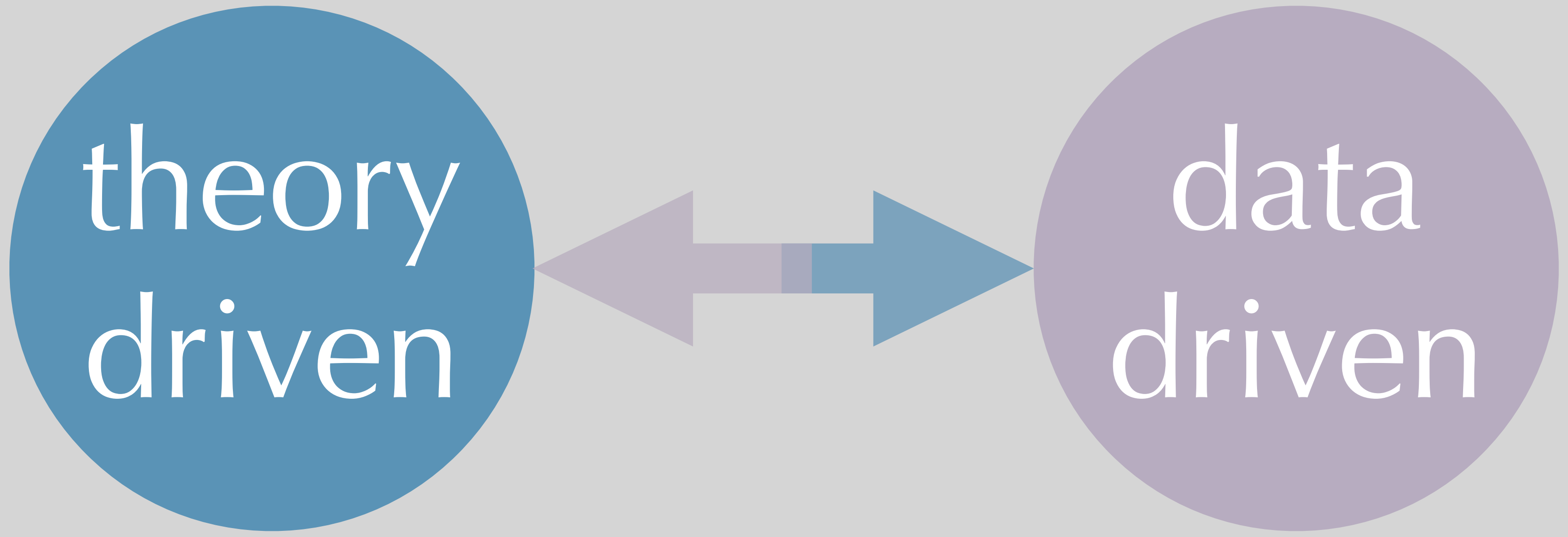


The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

out-of-sample predictive ability:

clear measure of effect size

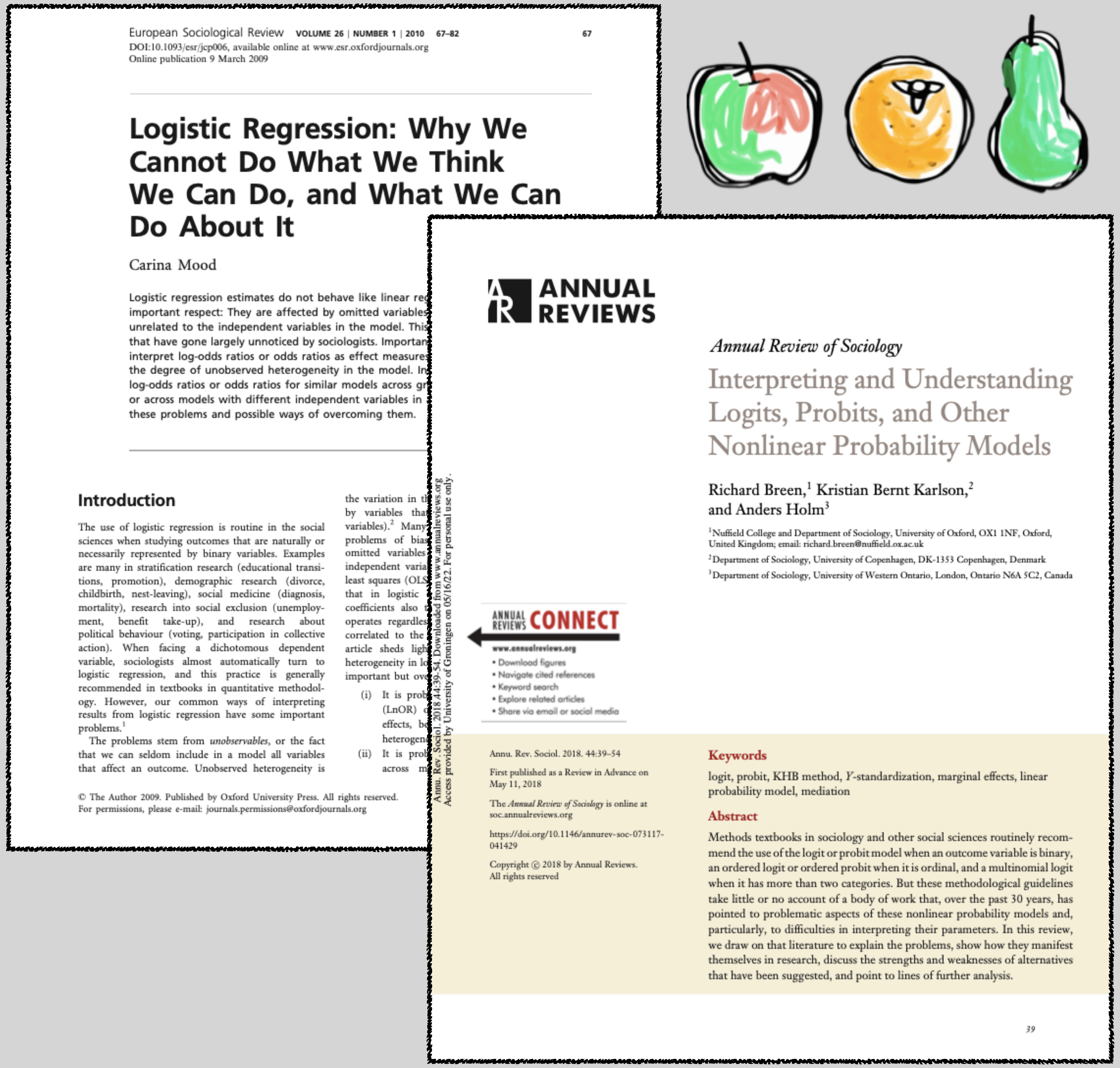


facilitates dialogue theory- and data-driven models

measure of distance theory and practice

out-of-sample predictive ability

- *is easy(ier) to understand*
- *can be compared across analytical techniques*
- *can be compared across models*
- *is less gameable*



European Sociological Review VOLUME 26 | NUMBER 1 | 2010 67–82 67
DOI:10.1093/esr/jcp006, available online at www.esr.oxfordjournals.org
Online publication 9 March 2009

Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It

Carina Mood

Logistic regression estimates do not behave like linear reg
important respect: They are affected by omitted variables
unrelated to the independent variables in the model. This
that have gone largely unnoticed by sociologists. Importan
interpret log-odds ratios or odds ratios as effect measures
the degree of unobserved heterogeneity in the model. In
log-odds ratios or odds ratios for similar models across gr
or across models with different independent variables in
these problems and possible ways of overcoming them.

Introduction

The use of logistic regression is routine in the social sciences when studying outcomes that are naturally or necessarily represented by binary variables. Examples are many in stratification research (educational transitions, promotion), demographic research (divorce, childbirth, nest-leaving), social medicine (diagnosis, mortality), research into social exclusion (unemployment, benefit take-up), and research about political behaviour (voting, participation in collective action). When facing a dichotomous dependent variable, sociologists almost automatically turn to logistic regression, and this practice is generally recommended in textbooks in quantitative methodology. However, our common ways of interpreting results from logistic regression have some important problems.[1]

The problems stem from *unobservables*, or the fact that we can seldom include in a model all variables that affect an outcome. Unobserved heterogeneity is

© The Author 2009. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

ANNUAL REVIEWS

Annual Review of Sociology

Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models

Richard Breen,[1] Kristian Bernt Karlson,[2] and Anders Holm[3]

[1]Nuffield College and Department of Sociology, University of Oxford, OX1 1NF, Oxford, United Kingdom; email: richard.breen@nuffield.ox.ac.uk
[2]Department of Sociology, University of Copenhagen, DK-1353 Copenhagen, Denmark
[3]Department of Sociology, University of Western Ontario, London, Ontario N6A 5C2, Canada

Annu. Rev. Sociol. 2018. 44:39–54. Downloaded from www.annualreviews.org
Access provided by University of Groningen on 05/16/22. For personal use only.

ANNUAL REVIEWS CONNECT
www.annualreviews.org
• Download figures
• Navigate cited references
• Keyword search
• Explore related articles
• Share via email or social media

Annu. Rev. Sociol. 2018. 44:39–54

First published as a Review in Advance on May 11, 2018

The *Annual Review of Sociology* is online at soc.annualreviews.org

https://doi.org/10.1146/annurev-soc-073117-041429

Copyright © 2018 by Annual Reviews. All rights reserved

Keywords

logit, probit, KHB method, *Y*-standardization, marginal effects, linear probability model, mediation

Abstract

Methods textbooks in sociology and other social sciences routinely recommend the use of the logit or probit model when an outcome variable is binary, an ordered logit or ordered probit when it is ordinal, and a multinomial logit when it has more than two categories. But these methodological guidelines take little or no account of a body of work that, over the past 30 years, has pointed to problematic aspects of these nonlinear probability models and, particularly, to difficulties in interpreting their parameters. In this review, we draw on that literature to explain the problems, show how they manifest themselves in research, discuss the strengths and weaknesses of alternatives that have been suggested, and point to lines of further analysis.

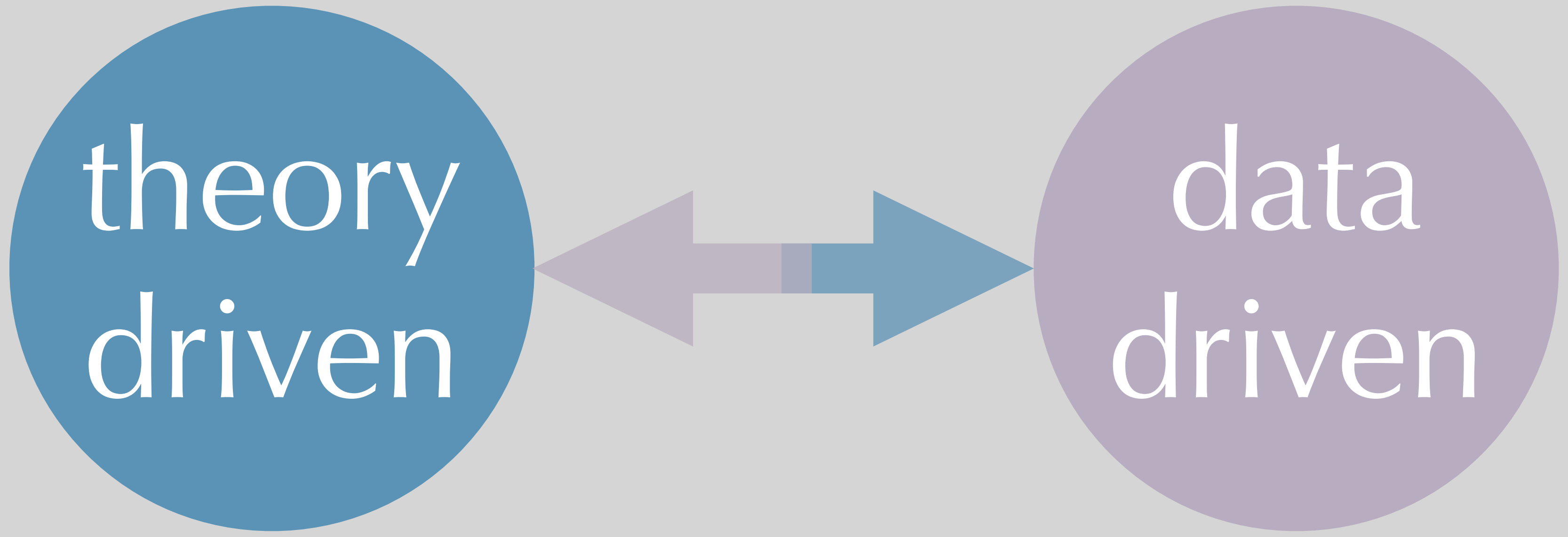
39

The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

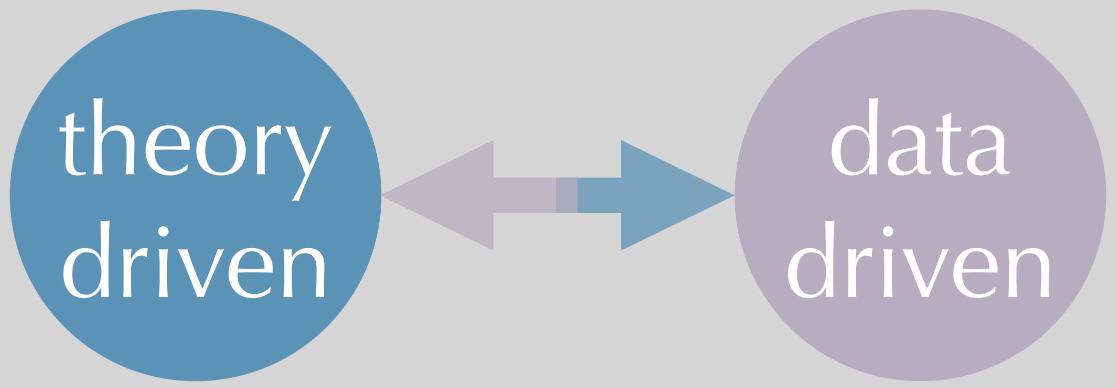
out-of-sample predictive ability:

clear measure of effect size



facilitates dialogue theory- and data-driven models

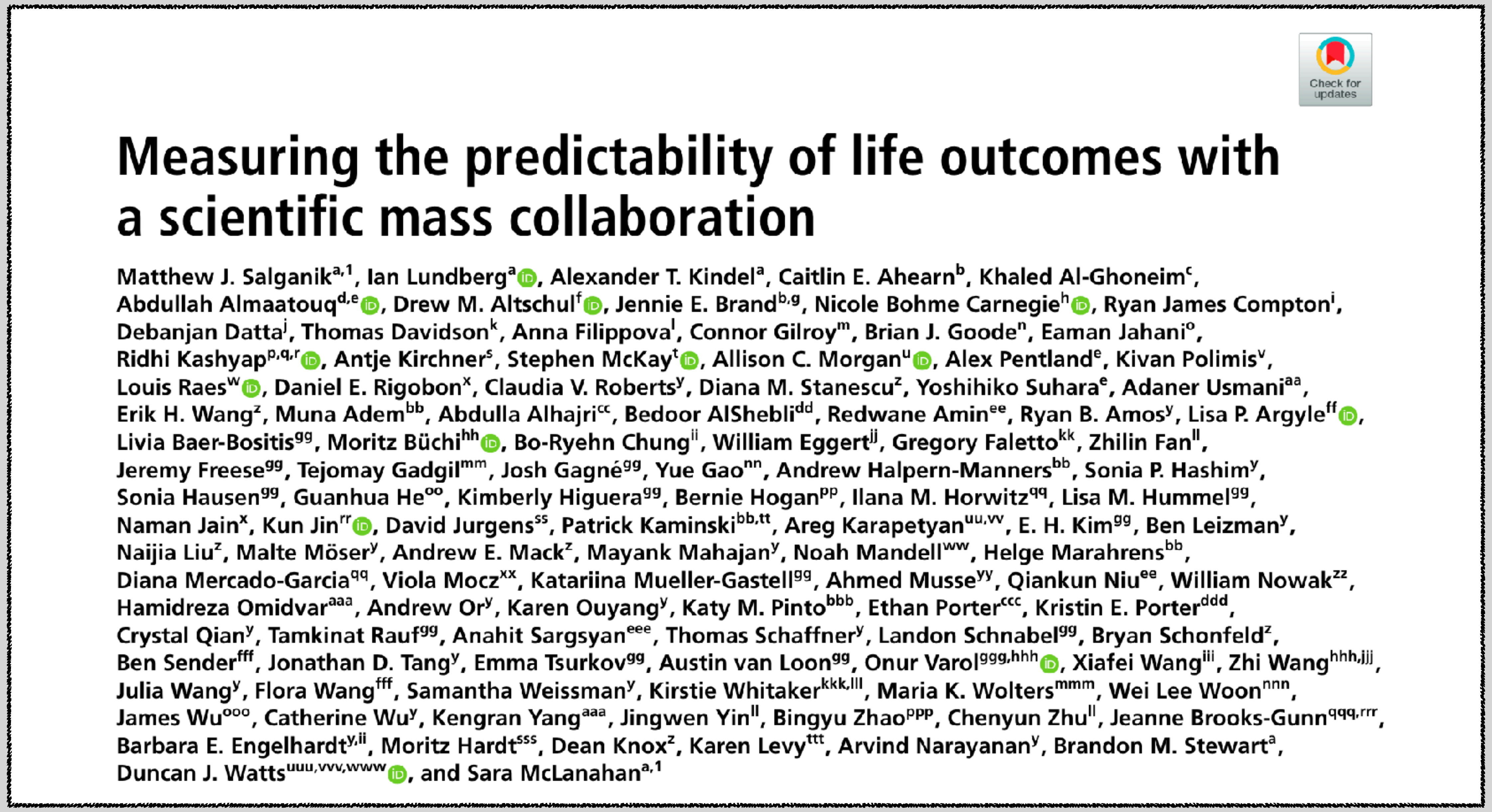
measure of distance theory and practice



theory-driven *vs* data-driven

theory-driven	data-driven
focus on (causal) estimates	*focus on predictive ability*
support based on p-value	*support based on prediction*
limited number of variables (k)	*k may be larger than n*
NHST weird theory-testing *long reign the linear model* *pet variable problem*	*estimates uninterpretable (sort of)* *computing intensive*

Predictability Crisis?

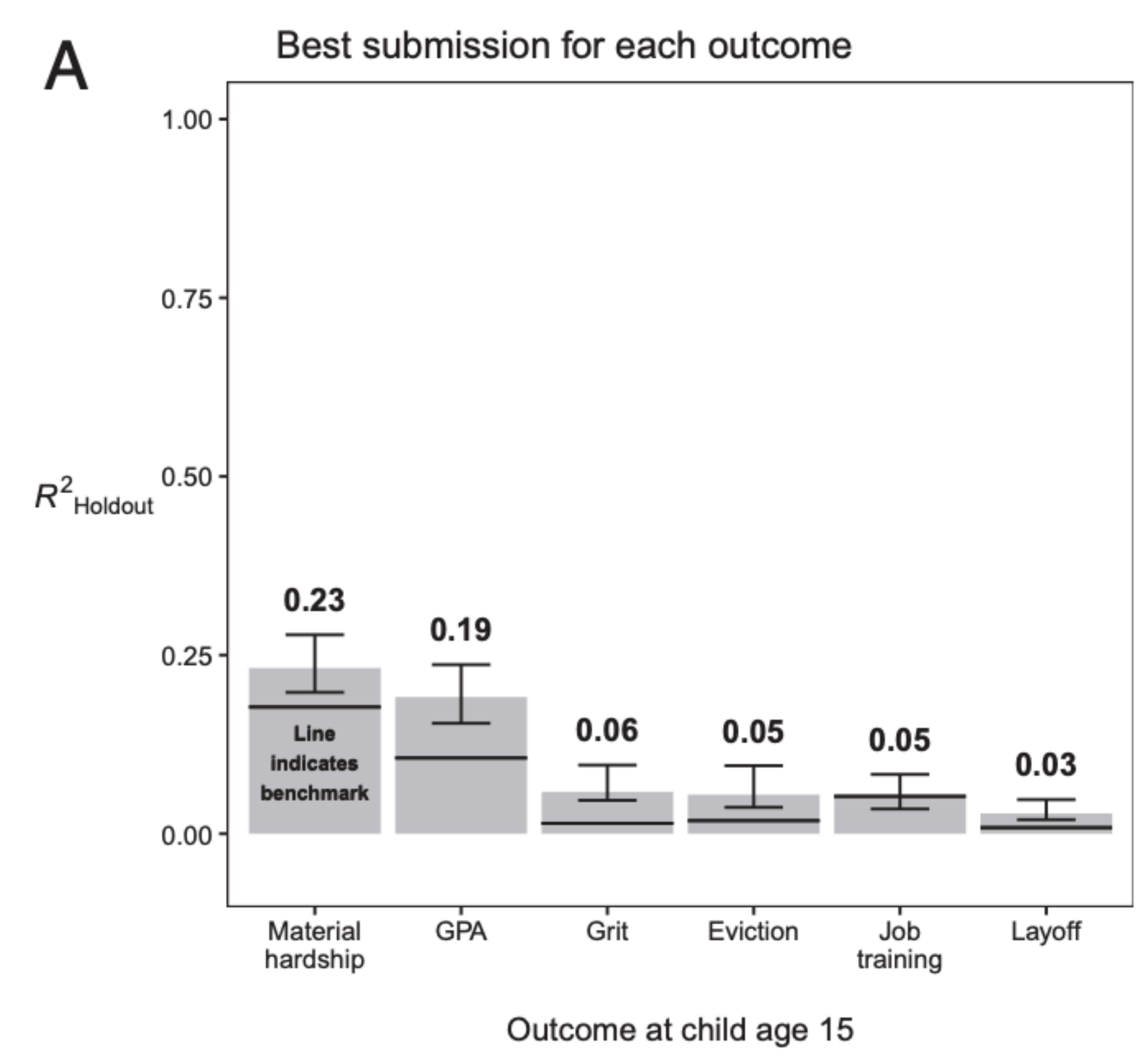


Check for updates

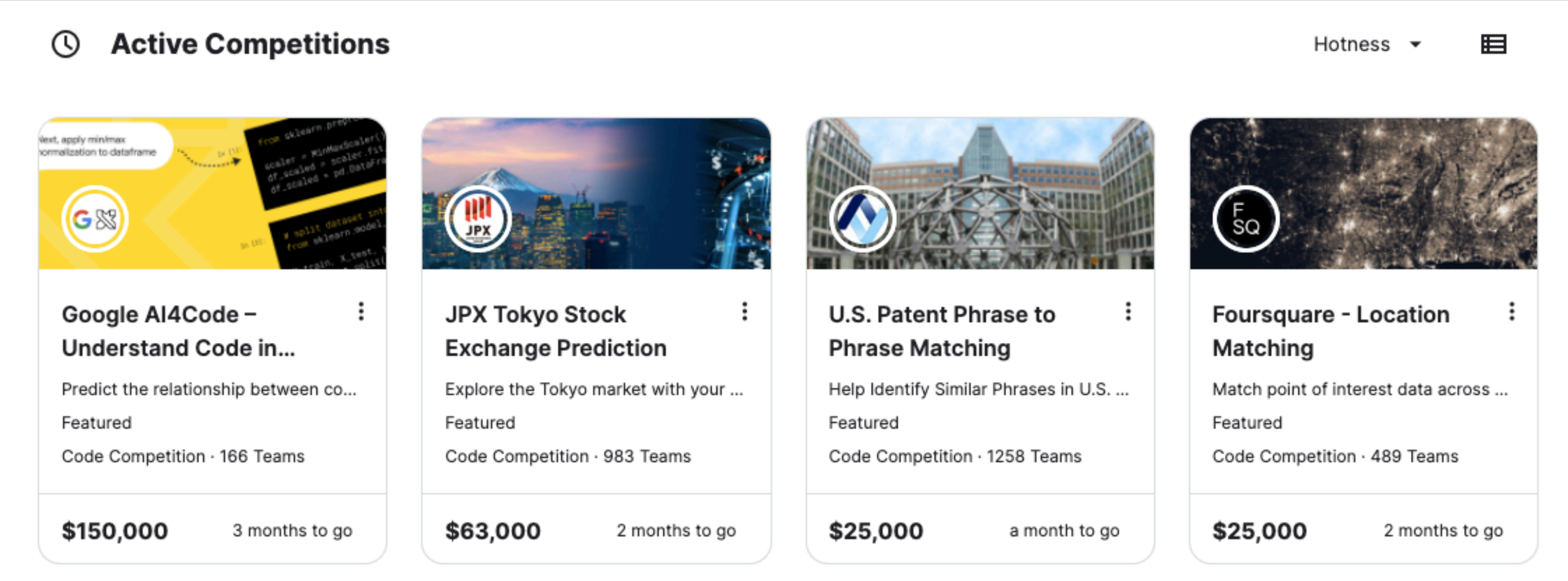
Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik[a,1], Ian Lundberg[a], Alexander T. Kindel[a], Caitlin E. Ahearn[b], Khaled Al-Ghoneim[c], Abdullah Almaatouq[d,e], Drew M. Altschul[f], Jennie E. Brand[b,g], Nicole Bohme Carnegie[h], Ryan James Compton[i], Debanjan Datta[j], Thomas Davidson[k], Anna Filippova[l], Connor Gilroy[m], Brian J. Goode[n], Eaman Jahani[o], Ridhi Kashyap[p,q,r], Antje Kirchner[s], Stephen McKay[t], Allison C. Morgan[u], Alex Pentland[e], Kivan Polimis[v], Louis Raes[w], Daniel E. Rigobon[x], Claudia V. Roberts[y], Diana M. Stanescu[z], Yoshihiko Suhara[e], Adaner Usmani[aa], Erik H. Wang[z], Muna Adem[bb], Abdulla Alhajri[cc], Bedoor AlShebli[dd], Redwane Amin[ee], Ryan B. Amos[y], Lisa P. Argyle[ff], Livia Baer-Bositis[gg], Moritz Büchi[hh], Bo-Ryehn Chung[ii], William Eggert[jj], Gregory Faletto[kk], Zhilin Fan[ll], Jeremy Freese[gg], Tejomay Gadgil[mm], Josh Gagné[gg], Yue Gao[nn], Andrew Halpern-Manners[bb], Sonia P. Hashim[y], Sonia Hausen[gg], Guanhua He[oo], Kimberly Higuera[gg], Bernie Hogan[pp], Ilana M. Horwitz[qq], Lisa M. Hummel[gg], Naman Jain[x], Kun Jin[rr], David Jurgens[ss], Patrick Kaminski[bb,tt], Areg Karapetyan[uu,vv], E. H. Kim[gg], Ben Leizman[y], Naijia Liu[z], Malte Möser[y], Andrew E. Mack[z], Mayank Mahajan[y], Noah Mandell[ww], Helge Marahrens[bb], Diana Mercado-Garcia[qq], Viola Mocz[xx], Katariina Mueller-Gastell[gg], Ahmed Musse[yy], Qiankun Niu[ee], William Nowak[zz], Hamidreza Omidvar[aaa], Andrew Or[y], Karen Ouyang[y], Katy M. Pinto[bbb], Ethan Porter[ccc], Kristin E. Porter[ddd], Crystal Qian[y], Tamkinat Rauf[gg], Anahit Sargsyan[eee], Thomas Schaffner[y], Landon Schnabel[gg], Bryan Schonfeld[z], Ben Sender[fff], Jonathan D. Tang[y], Emma Tsurkov[gg], Austin van Loon[gg], Onur Varol[ggg,hhh], Xiafei Wang[iii], Zhi Wang[hhh,jjj], Julia Wang[y], Flora Wang[fff], Samantha Weissman[y], Kirstie Whitaker[kkk,lll], Maria K. Wolters[mmm], Wei Lee Woon[nnn], James Wu[ooo], Catherine Wu[y], Kengran Yang[aaa], Jingwen Yin[ll], Bingyu Zhao[ppp], Chenyun Zhu[ll], Jeanne Brooks-Gunn[qqq,rrr], Barbara E. Engelhardt[y,ii], Moritz Hardt[sss], Dean Knox[z], Karen Levy[ttt], Arvind Narayanan[y], Brandon M. Stewart[a], Duncan J. Watts[uuu,vvv,www], and Sara McLanahan[a,1]

data challenge:
predicting life outcomes
based on ~6000 variables
by 160 teams
both theory- & data-driven



kaggle



“secret sauce of data science

Donoho, 2015

data
challenge

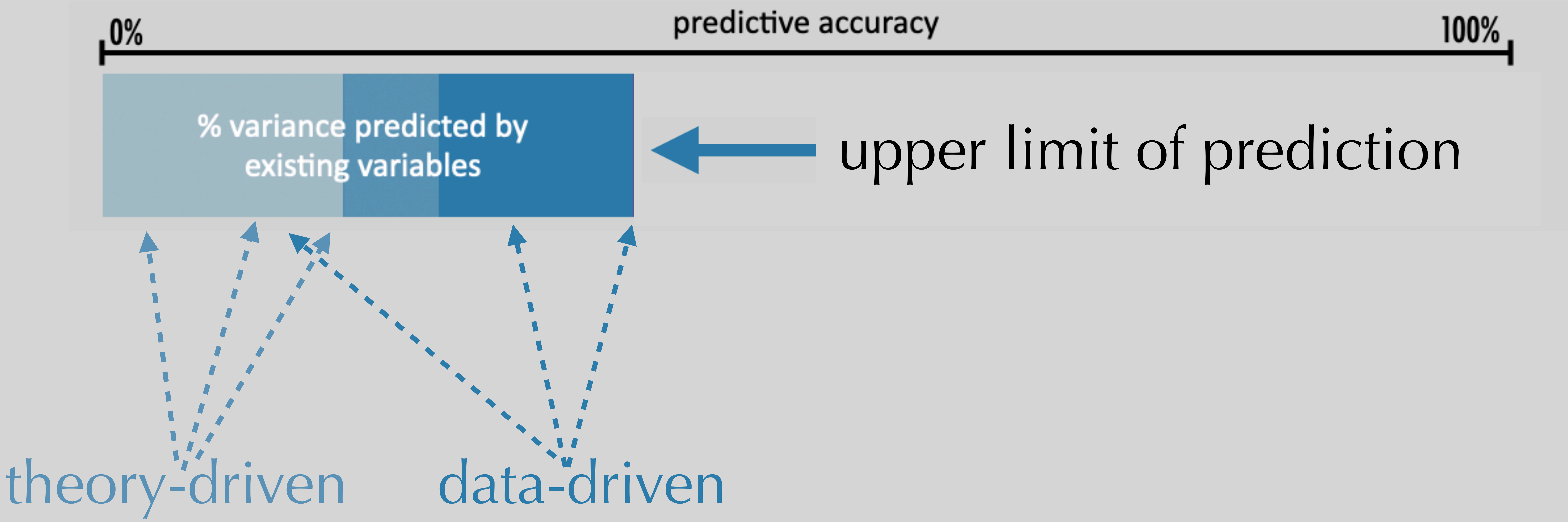
theory driven

data driven

theory- and data-driven teams
engage in common task
using common data
and common metric

Data Challenge

theory- and data-driven teams
engage in common task
using common data
and common metric



Prediction Benchmarks

baseline benchmarks	upper limit benchmarks
established with state-of-the-art theory	*established with state-of-the-art statistical learning tools*

Prediction Benchmarks

baseline benchmarks	upper limit benchmarks
established with state-of-the-art theory	*established with state-of-the-art statistical learning tools*

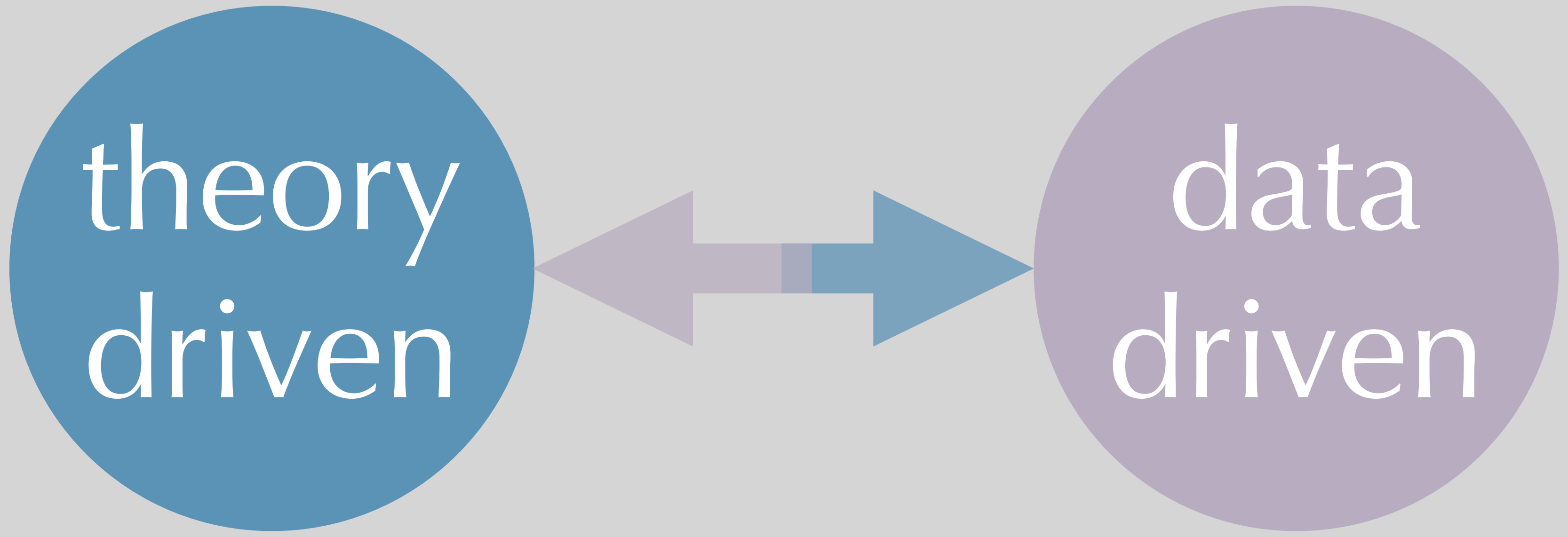
> Progress usually comes from many small improvements; a change of 1% can be a reason to break out the champagne
>
> Liberman, 2012

The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

out-of-sample predictive ability:

clear measure of effect size



facilitates dialogue theory- and data-driven models

measure of distance theory and practice

out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

Articles

journal of peace RESEARCH

Journal of Peace Research
47(4) 363–375

DOI: 10.1177/0022343309356491

The perils of policy by p-value: Predicting civil conflicts

Michael D Ward
Department of Political Science, Duke University
Brian D Greenhill
Department of Political Science, University of Washington
Kristin M Bakke
Department of Political Science, University College London

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
1–11

DOI: 10.1177/1745691620969650
www.psychologicalscience.org/PPS
SAGE

So Useful as a Good Theory? The Practicality Crisis in (Social) Psychological Theory

Elliot T. Berkman and Sylas M. Wilson
Department of Psychology and Center for Translational Neuroscience, University of Oregon

out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

PNAS

Why significant variables aren't automatically good predictors

Adeline Lo[a], Herman Chernoff[b,1], Tian Zheng[c], and Shaw-Hwa Lo[c,1]

[a]Department of Political Science, University of California, San Diego, La Jolla, CA 92093; [b]Department of Statistics, Harvard University, Cambridge, MA 02138; and [c]Department of Statistics, Columbia University, New York, NY 10027

Contributed by Herman Chernoff, September 17, 2015 (sent for review December 15, 2014)

Thus far, genome-wide association studies (GWAS) have been disappointing in the inability of investigators to use the results of

From the scientist's point of view there are two basic problems, complicated by the large size of the data set. These are variable

out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

> "Social scientists studying the life course must find a way to reconcile a widespread belief that understanding has been generated by these data—as demonstrated by more than 750 published journal articles using the Fragile Families data with the fact that the very same data could not yield accurate predictions of these important outcomes.

The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

microsimulation can advance traditional statistical modelling

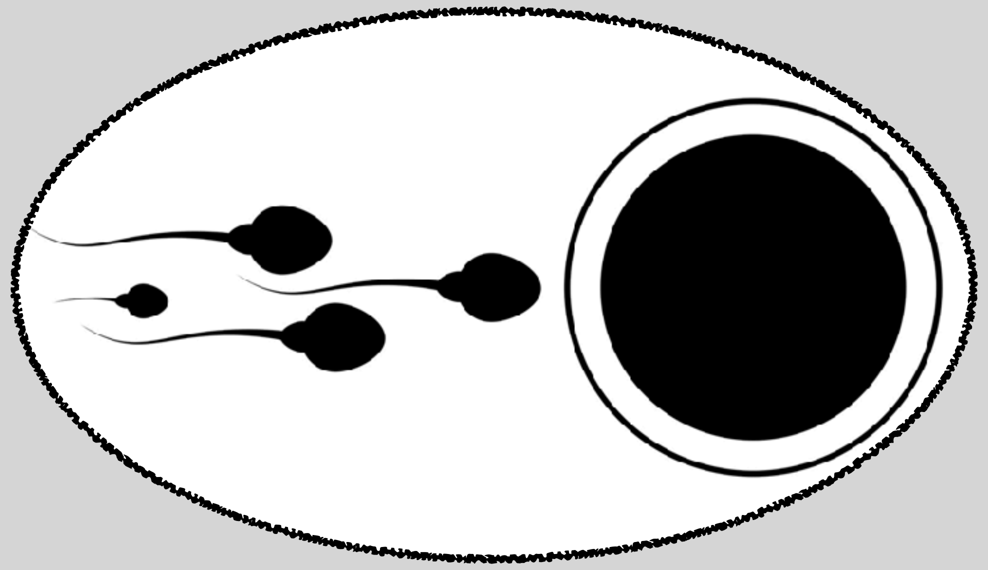
The Proposal

microsimulation can
advance traditional
statistical modelling

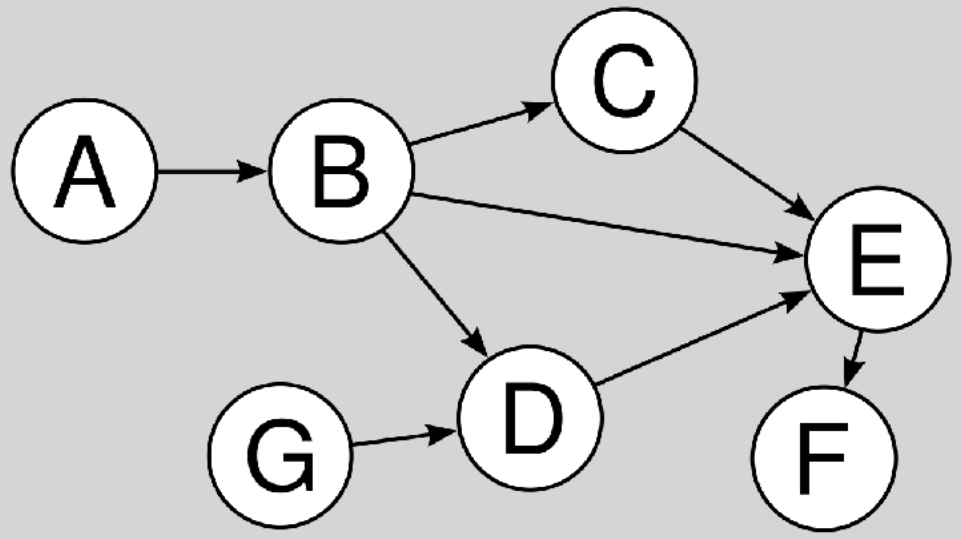
The Proposal

microsimulation can advance traditional statistical modelling

microsimulation can:

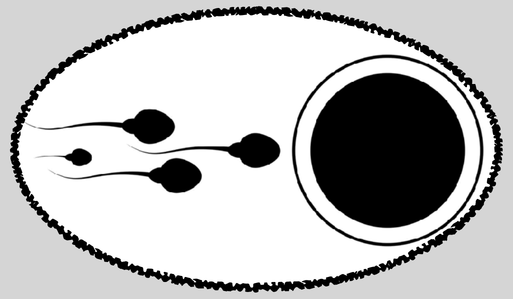


include biological information

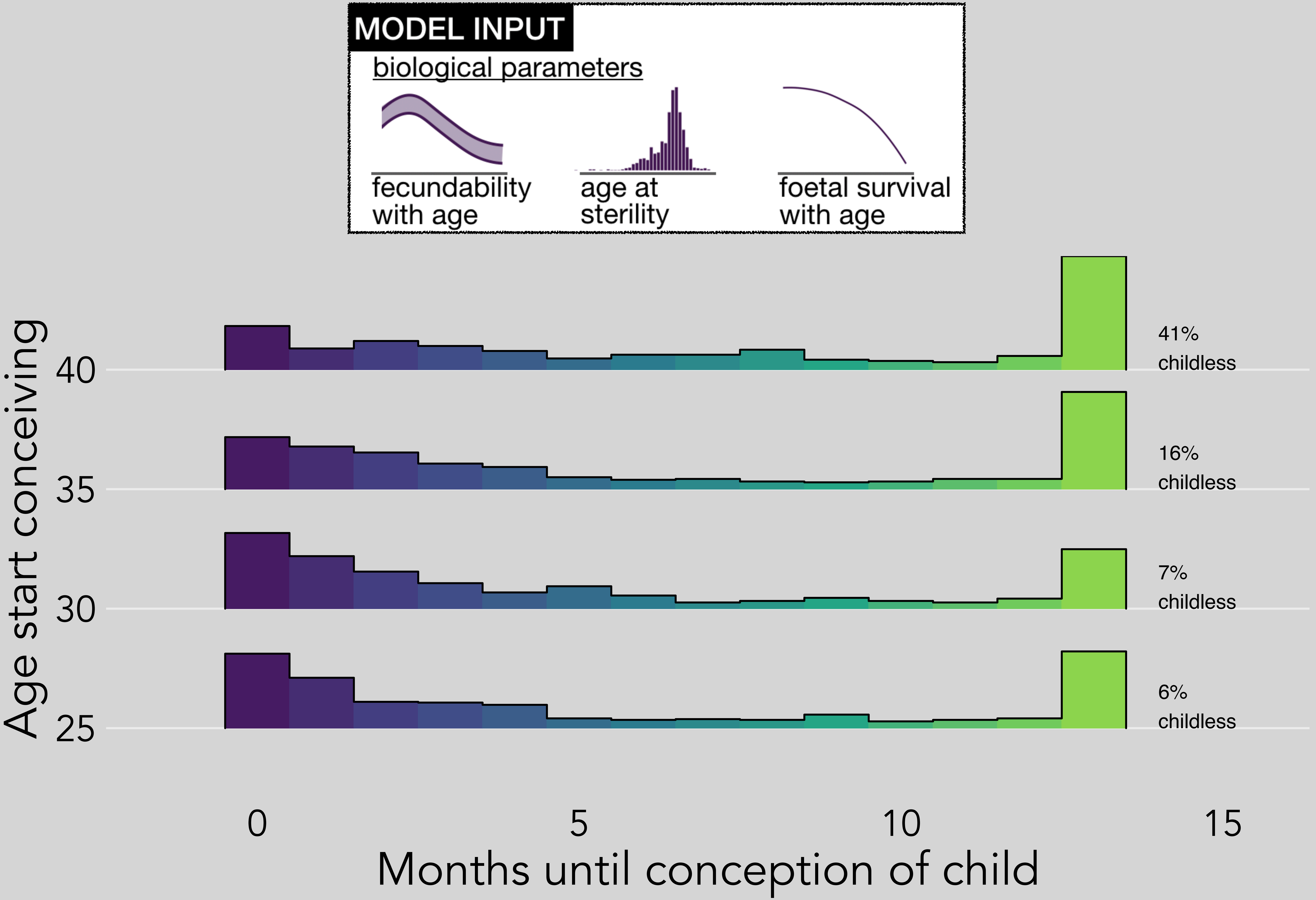


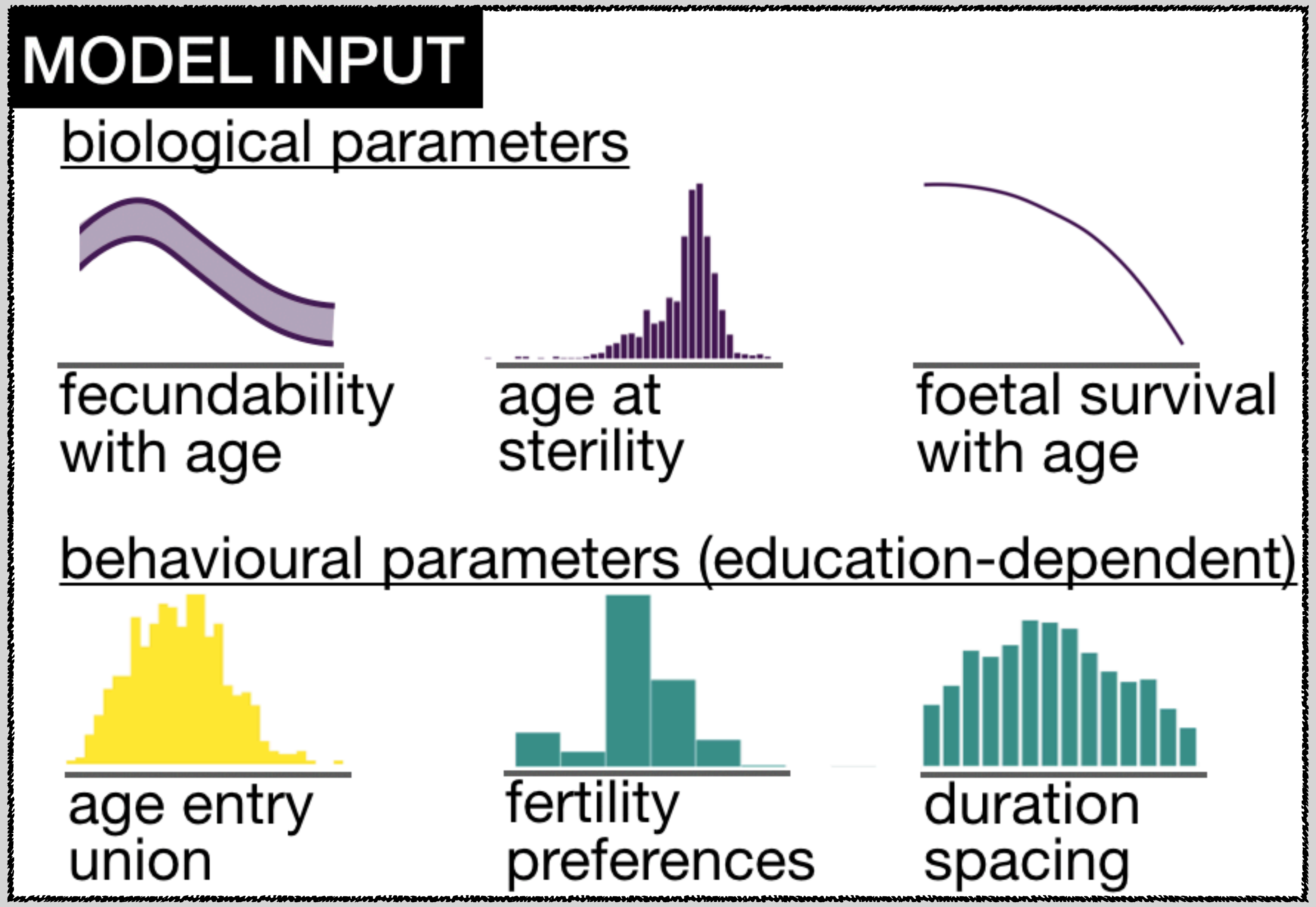
test (causal) mechanisms

quantify unpredictability




MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age


MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age
Age start conceiving
40
35
30
25
41% childless
16% childless
7% childless
6% childless
0
5
10
15
Months until conception of child



determines whether and when people would like to conceive

determines whether and when people conceive

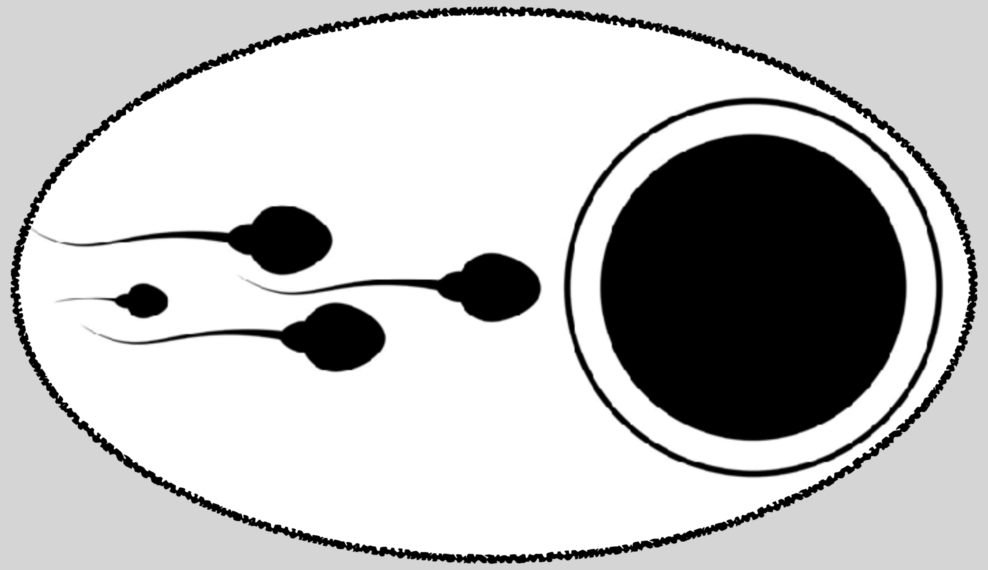

MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age
behavioural parameters (education-dependent)
age entry union
fertility preferences
duration spacing
MODEL RUN
Randomly determined traits individual 1
in union =22 | spac. =5 | pref. =2 | fecund. =0.3 | steril. =43 | edu. =high
stochasticity
stochasticity
1
22 union
27 start conceiving
29 birth
34 start conceiving
38 miscarriage
43 sterility
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low
stochasticity
10000
26 union
27 start conceiving
31 birth
no further children preferred


MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age
behavioural parameters (education-dependent)
age entry union
fertility preferences
duration spacing
MODEL RUN
Randomly determined traits individual 1
in union =22 | spac. =5 | pref. =2 | fecund. =0.3 | steril. =43 | edu. =high
stochasticity
stochasticity
22 union
27 start conceiving
29 birth
34 start conceiving
38 miscarriage
43 sterility
1
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low
stochasticity
26 union
27 start conceiving
31 birth
no further children preferred
10000
MODEL OUTPUT
due to:
partner status
preferences
stochasticity
% childless
low high
education
frequency
age first birth
frequency
children
'TRUTH'
% childless
low high
education
frequency
age first birth
frequency
children

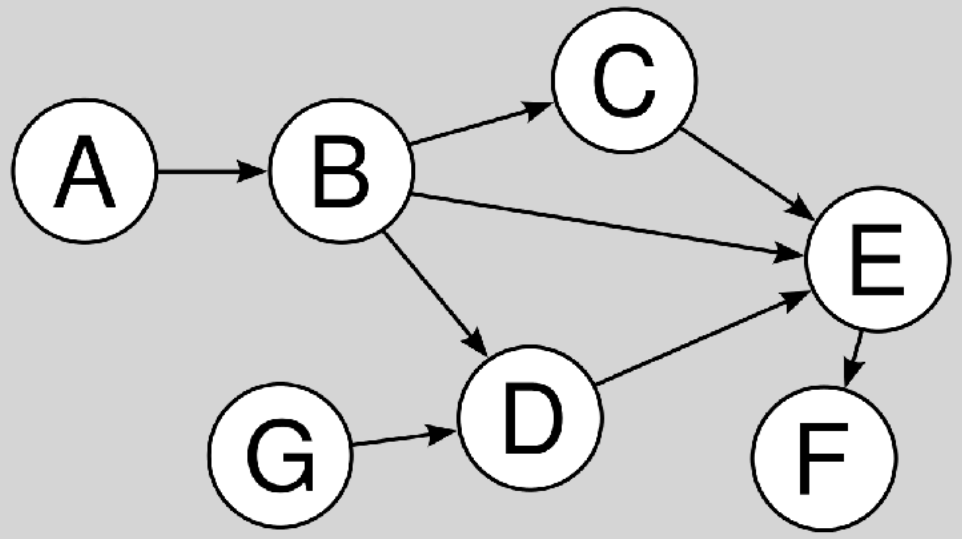
The Proposal

microsimulation can advance traditional statistical modelling

microsimulation can:

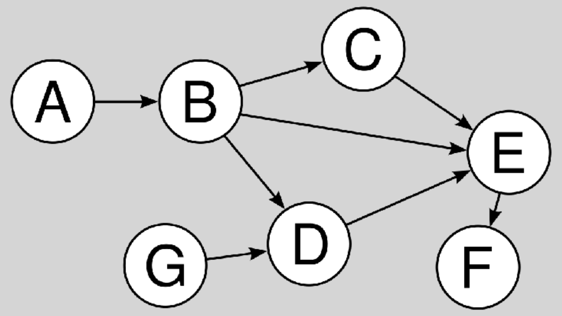


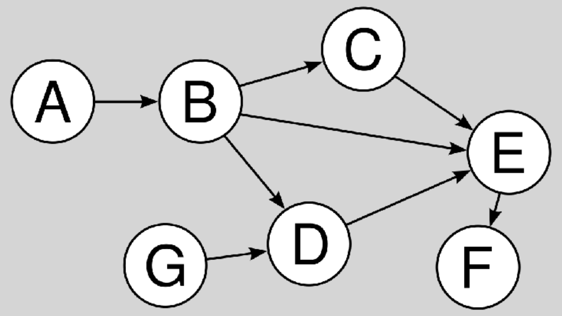
include biological information

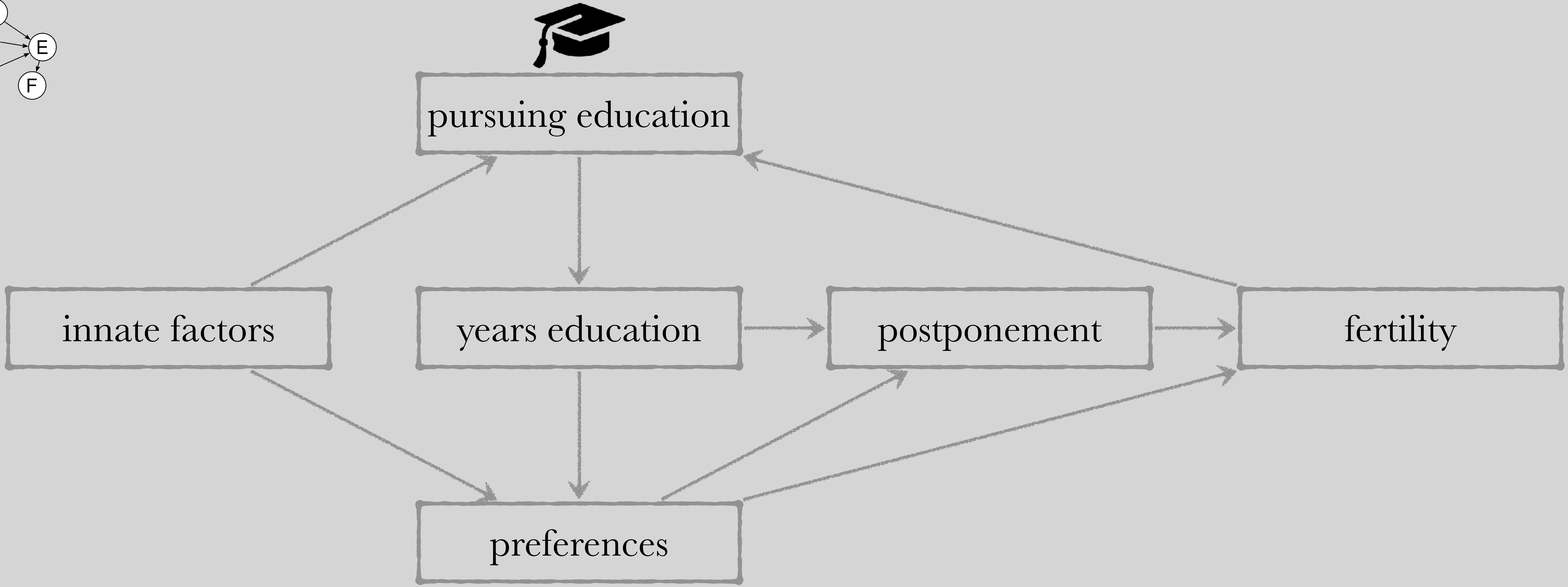


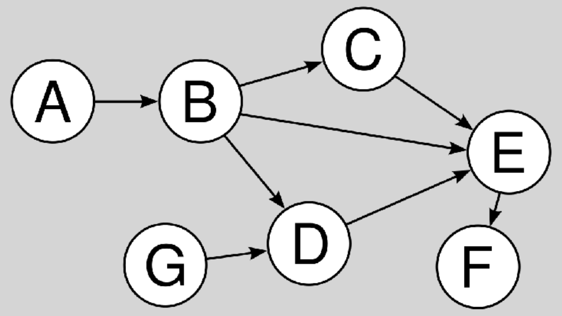
test (causal) mechanisms

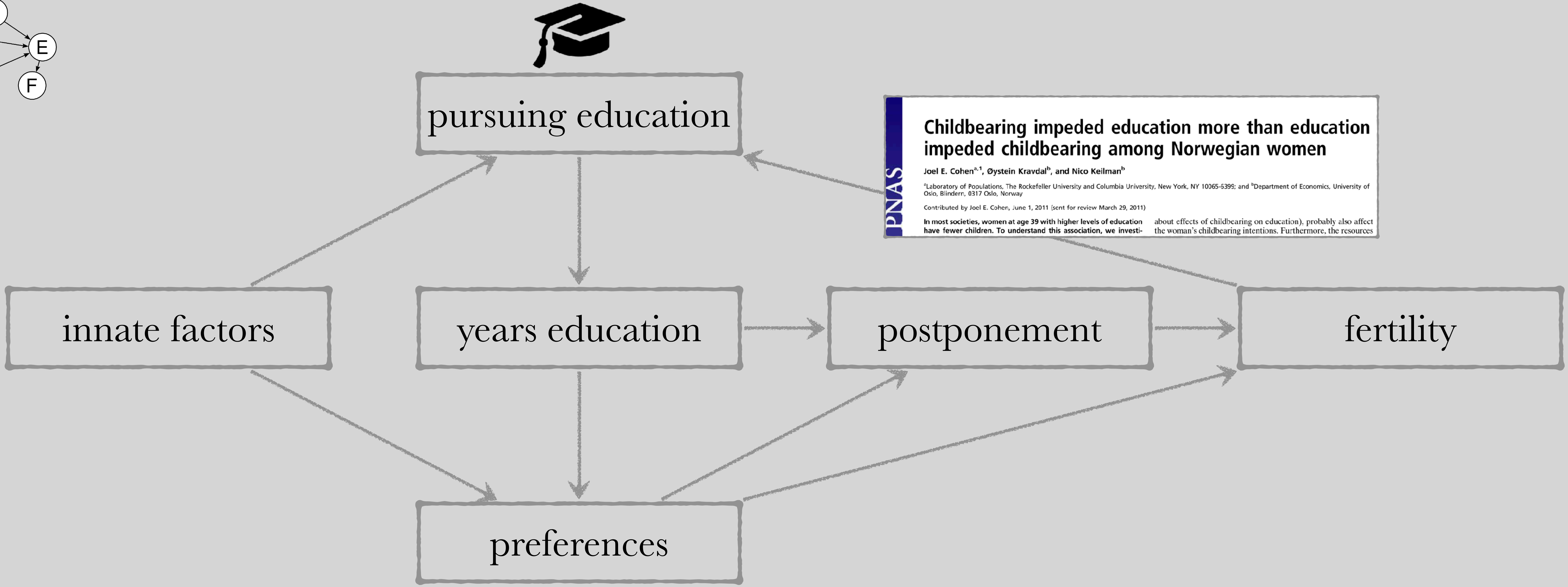
quantify unpredictability

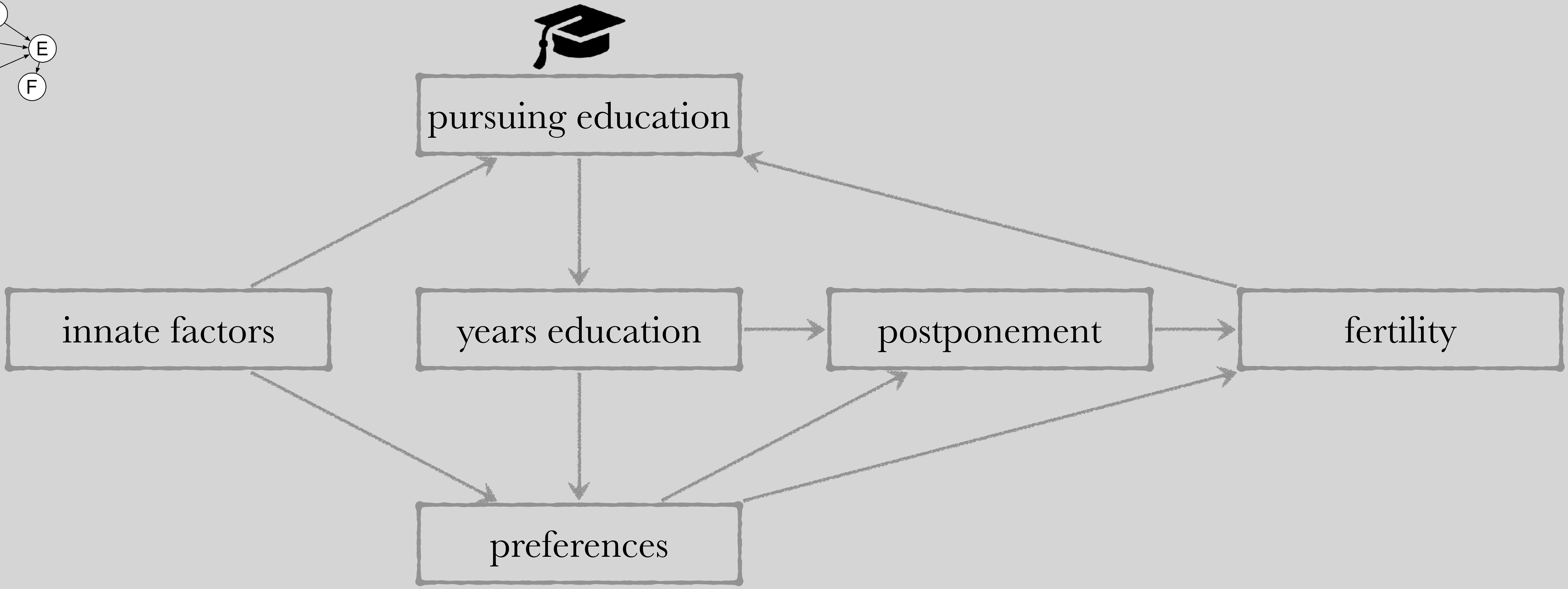

A
B
C
D
E
F
G


A
B
C
D
E
F
G

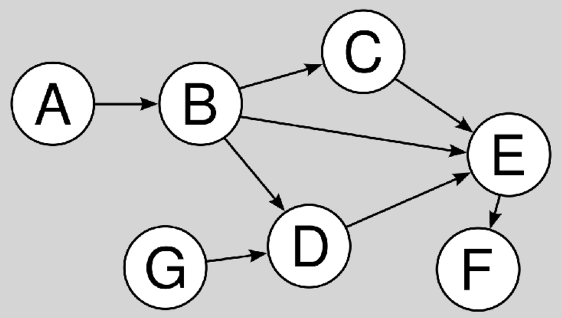

pursuing education
innate factors
years education
postponement
fertility
preferences

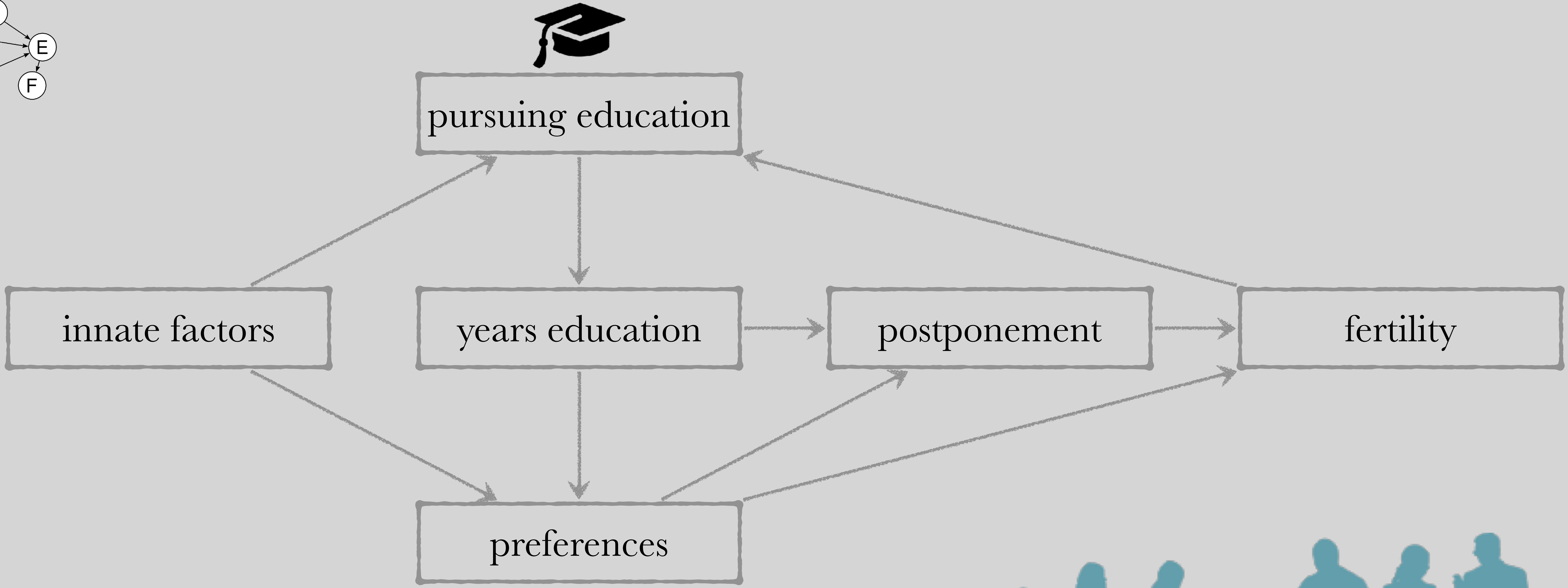

A
B
C
D
E
F
G

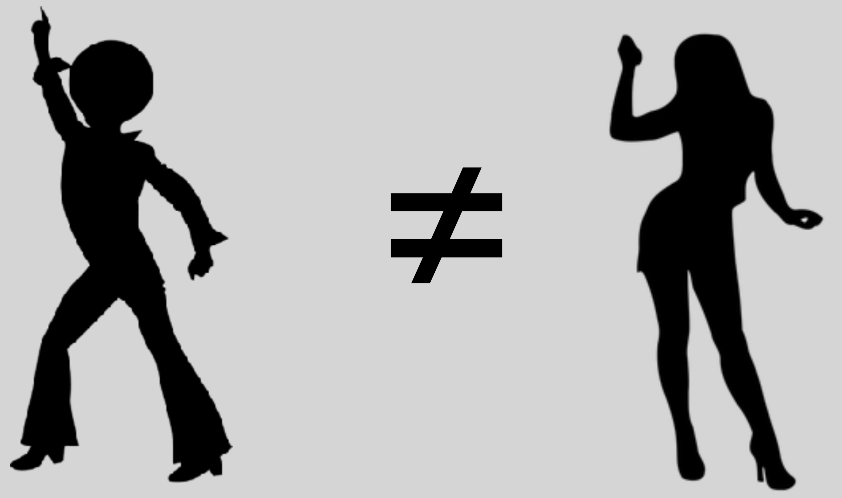

pursuing education
innate factors
years education
postponement
fertility
preferences
PNAS
Childbearing impeded education more than education impeded childbearing among Norwegian women
Joel E. Cohen[a,1], Øystein Kravdal[b], and Nico Keilman[b]
[a]Laboratory of Populations, The Rockefeller University and Columbia University, New York, NY 10065-6399; and [b]Department of Economics, University of Oslo, Blindern, 0317 Oslo, Norway
Contributed by Joel E. Cohen, June 1, 2011 (sent for review March 29, 2011)
In most societies, women at age 39 with higher levels of education have fewer children. To understand this association, we investi-
about effects of childbearing on education), probably also affect the woman's childbearing intentions. Furthermore, the resources



What Kind of Data Would We need to Address This Model?





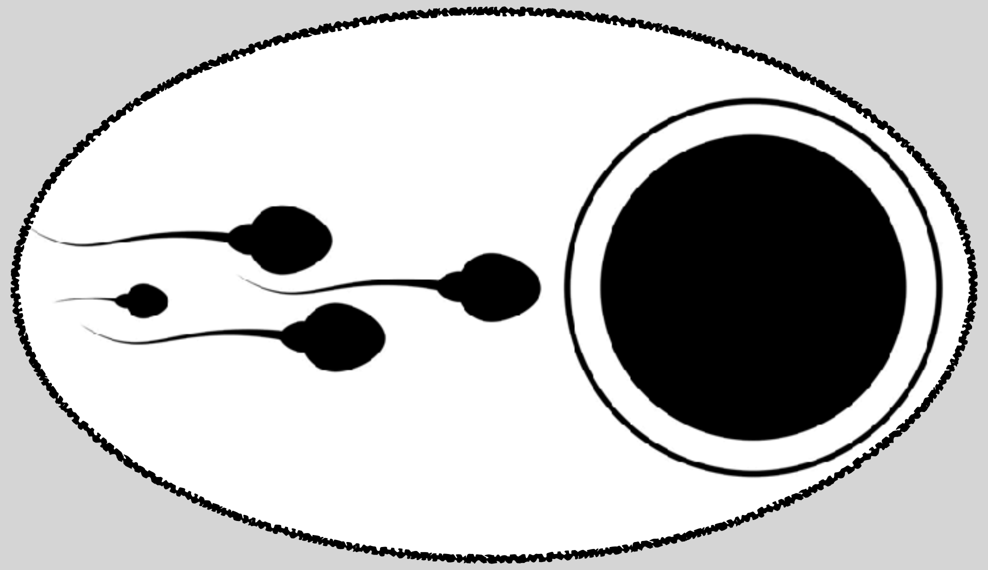


What Kind of Data Would We need to Address This Model?

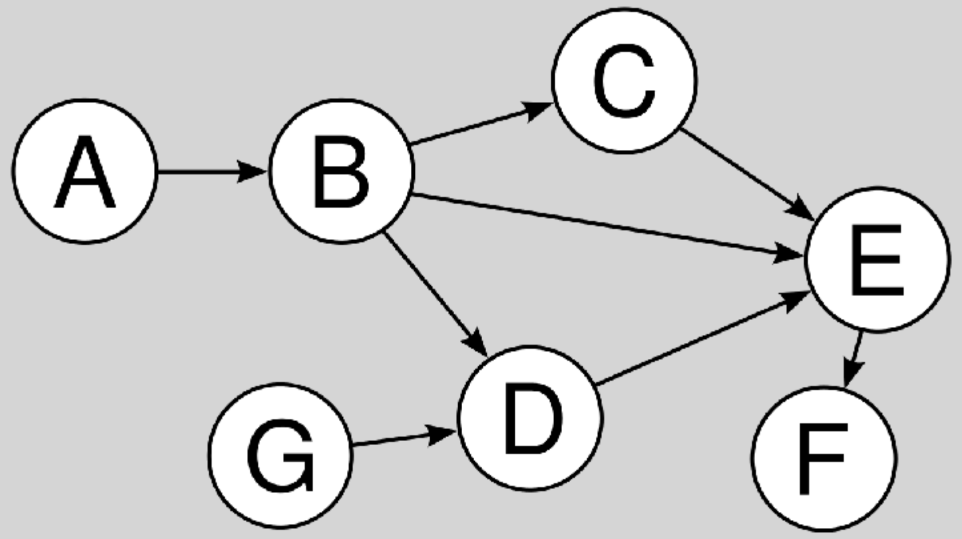
The Proposal

microsimulation can
advance traditional
statistical modelling

microsimulation can:



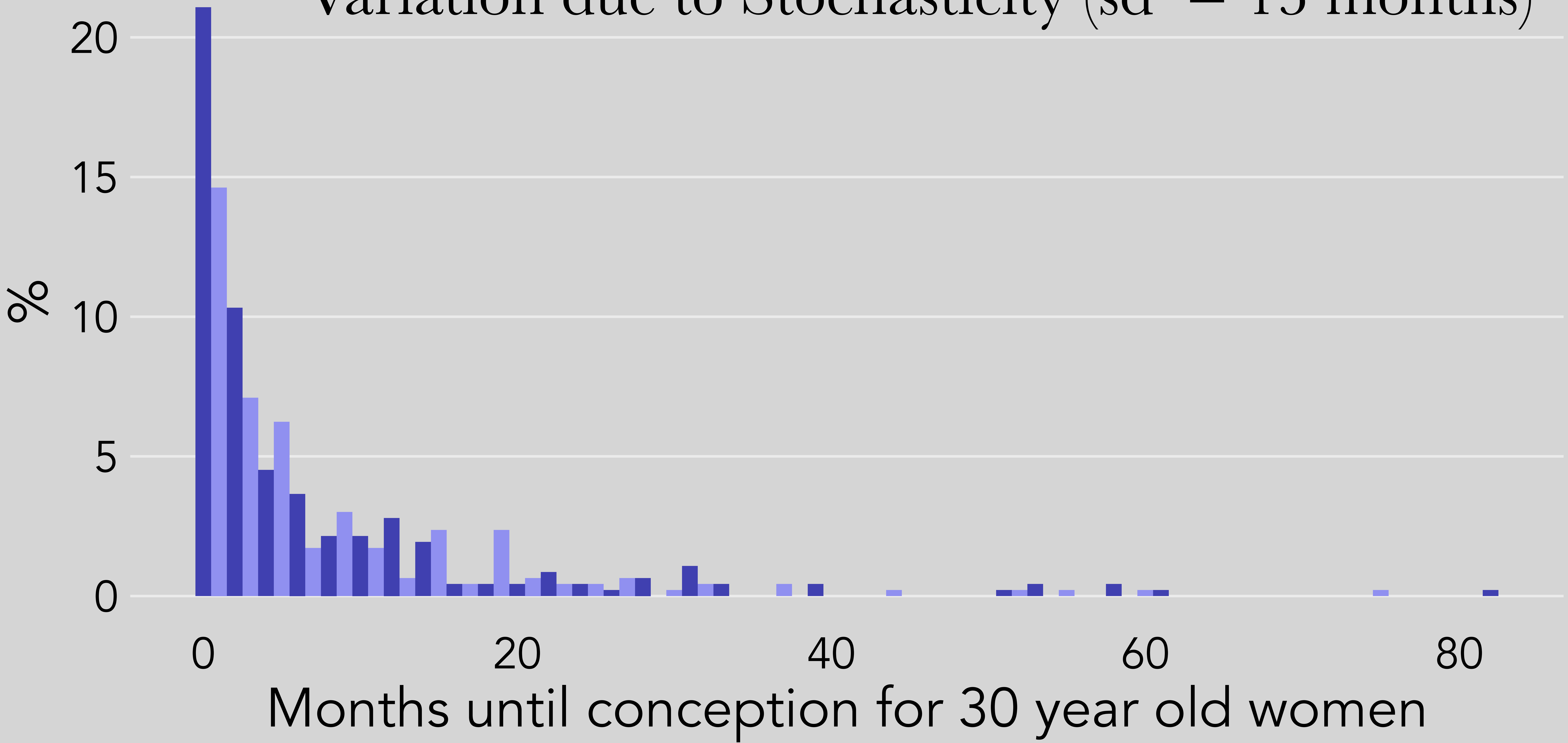
include
biological
information



test (causal)
mechanisms

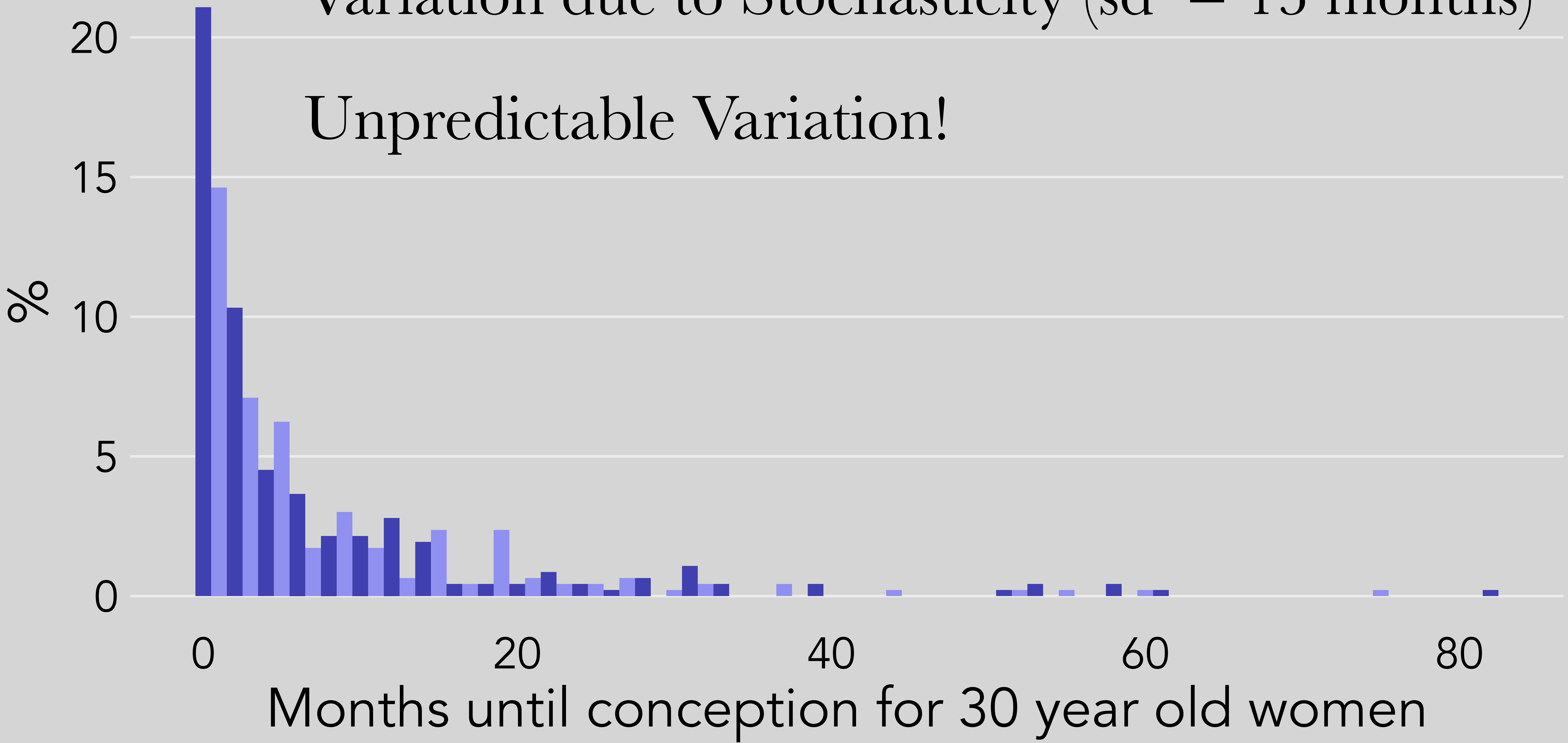
quantify
unpredictability

Variation due to Stochasticity (sd = 13 months)



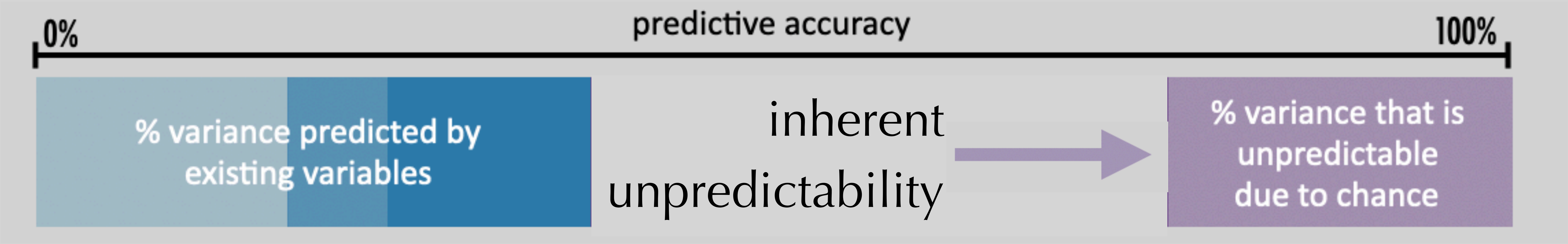
Variation due to Stochasticity (sd = 13 months)

Unpredictable Variation!

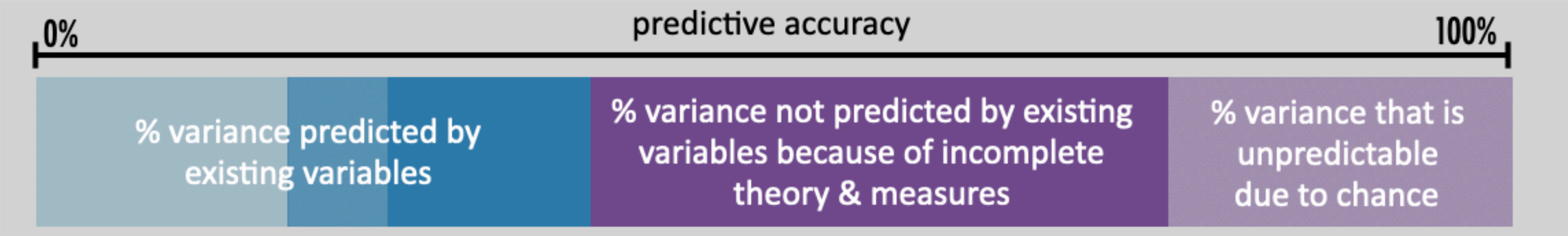


MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age
behavioural parameters (education-dependent)
age entry union
fertility preferences
duration spacing
MODEL RUN
Randomly determined traits individual 1
in union =22 | spac. =5 | pref. =2 | fecund. =0.3 | steril. =43 | edu. =high
stochasticity
stochasticity
22 union
27 start conceiving
29 birth
34 start conceiving
38 miscarriage
43 sterility
1
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low
stochasticity
26 union
27 start conceiving
31 birth
no further children preferred
10000
MODEL OUTPUT
due to:
partner status
preferences
stochasticity
% childless
low high
education
frequency
age first birth
frequency
children
'TRUTH'
% childless
low high
education
frequency
age first birth
frequency
children

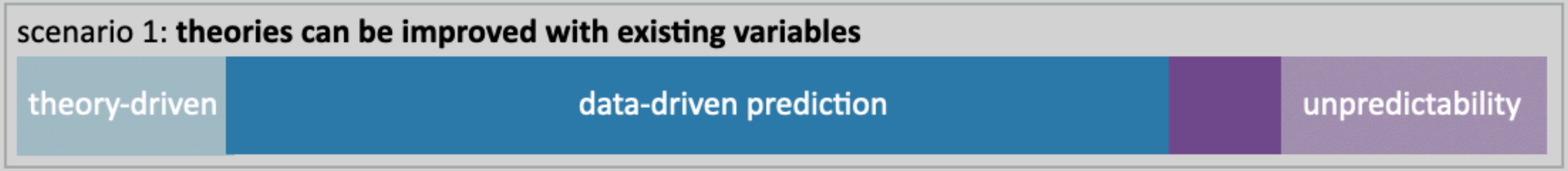
Unpredictable Variation



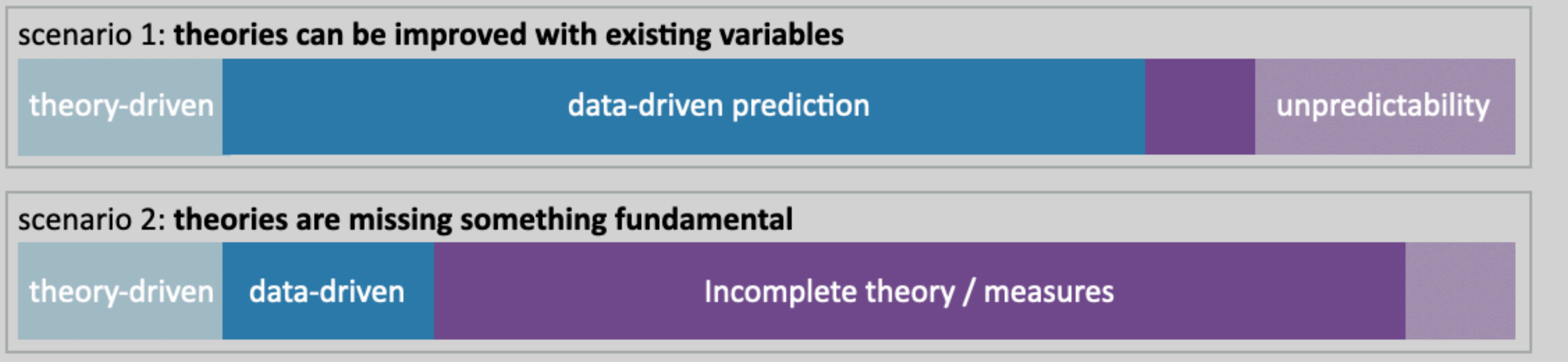
Unique Insight into State of Field



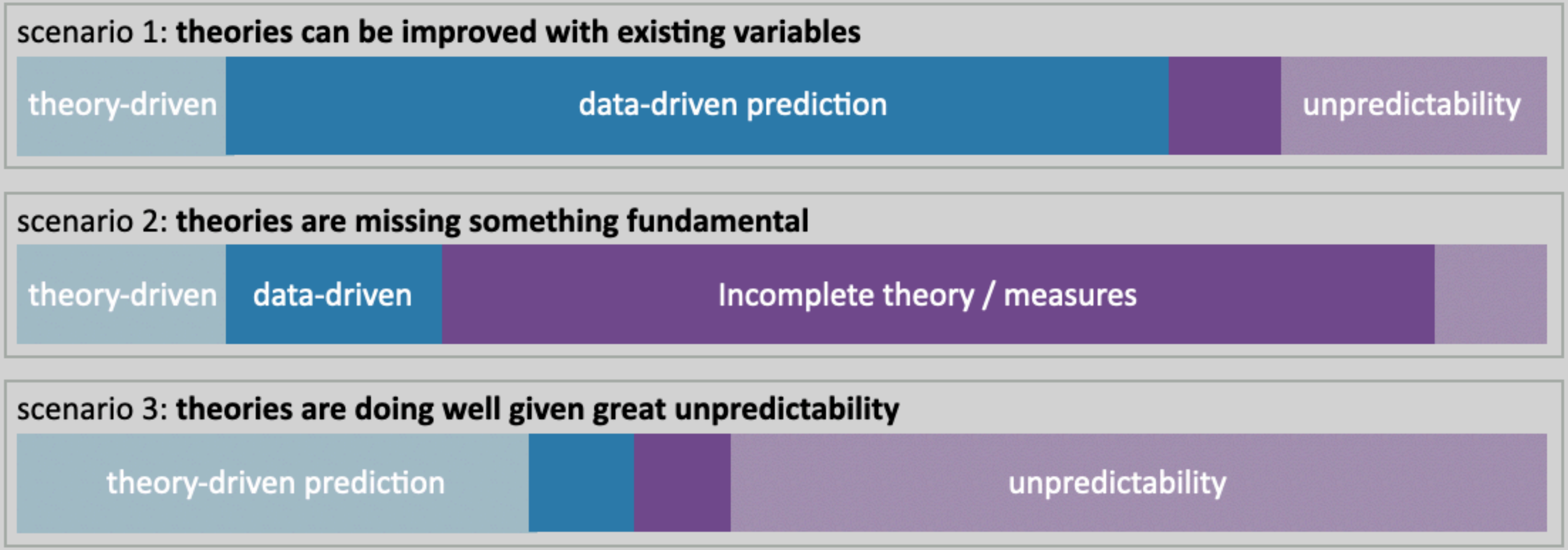
Unique Insight into State of Field



Unique Insight into State of Field



Unique Insight into State of Field



The Proposal

a shift towards **prediction** leads to a more reliable and useful social science

microsimulation can advance traditional statistical modelling

This mess we're in?

Or how simulation and prediction will advance (demographic) research

gert stulp — gertstulp.com — g.stulp@rug.nl