PSYCH 201B

Statistical Intuitions for Social Scientists

What is statistics?

01/08/2025

Today's Plan

- Background survey feedback and course adjustments
- What is statistics? What are the fundamentals?
  - Mini-discussion
- The two cultures of statistical modeling
  - Mini-discussion
- Bias, variance & ecological validity
  - Mini-discussion

About you!

Programming in Python

14 responses

Familiarity using git/github

14 responses

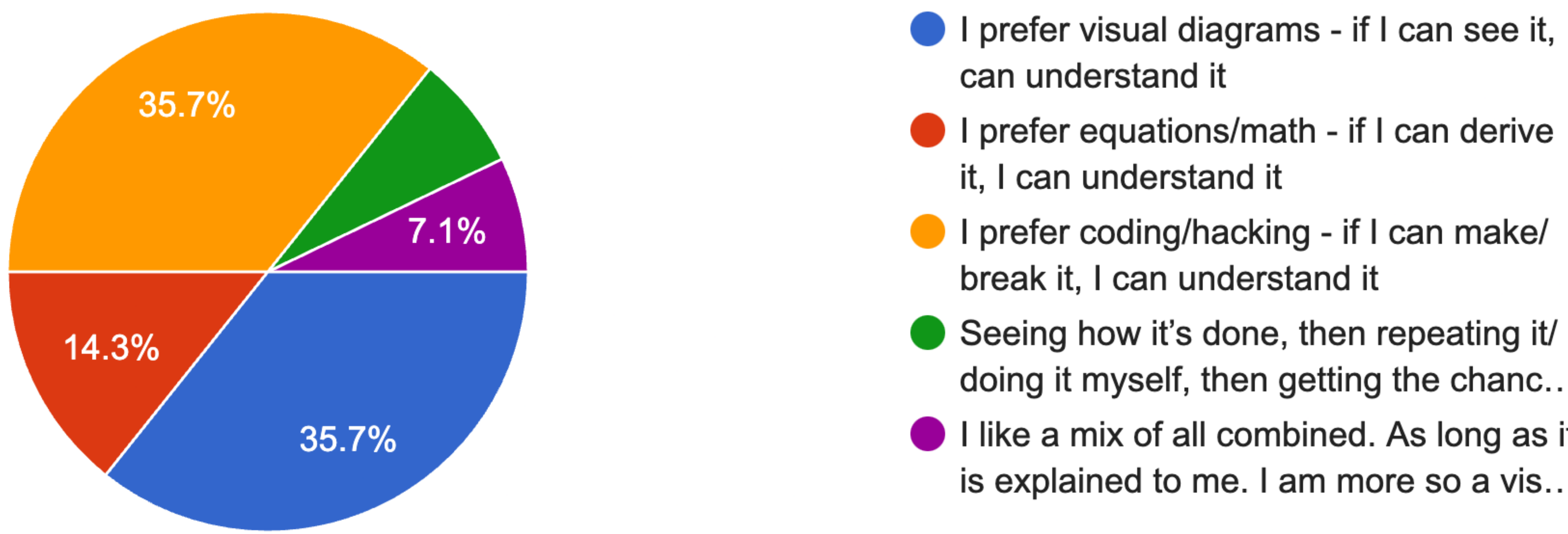
How comfortable are you with linear algebra?

14 responses

We’ll try to stick to visuals + code!

Given your current experience, which approach to learning statistical concepts feels most intuitive to you?

14 responses



Things you're **excited** about

- Python
- Regression models and applications
- Understanding statistical principles

Things you're **worried** about

- Python (and coding in general)
- Regression models and applications
- Understanding statistical principles

How we'll learn

- Instead of classic lecture/lab format we'll use **lecture** time to do **interactive coding too**
- For the next week or so we'll **assign tutorial notebooks** in addition to readings
  - Focus on the notebooks
  - Fine to skim readings
- Goal: get you more comfortable with Python while introducing high-level ideas
- Result: we can use Python as a "bicycle-for-the-mind" as we continue learning

*I think one of the things that really separates us from the high primates is that we're **tool builders**. I read a study that measured the efficiency of locomotion for various species on the planet. The condor used the least energy to move a kilometer. And, humans came in with a rather unimpressive showing, about a third of the way down the list. It was not too proud a showing for the crown of creation.*

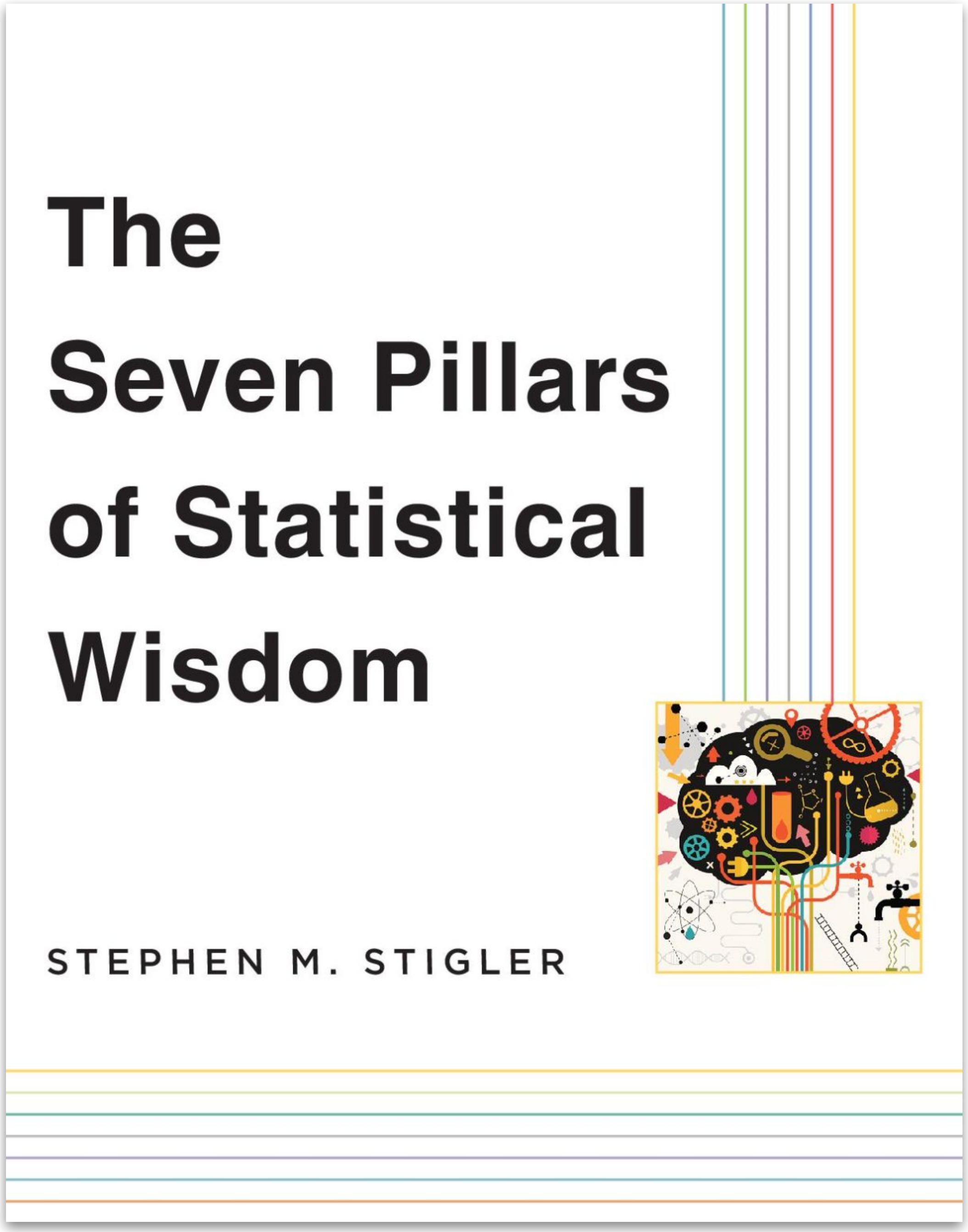
*So, that didn't look so good. But, then somebody at Scientific American had the insight to test the efficiency of **locomotion for a man on a bicycle**. And, a man on a bicycle, a human on a bicycle, blew the condor away, completely off the top of the charts.*

*And that's what a computer is to me. What **a computer** is to me **is** it's the **most remarkable tool that we've ever come up** with, and **it's the equivalent of a bicycle for our minds**."*

<u>~ Steve Jobs, 1995</u>

What is statistics?

What is statistics?



Fundamental concepts

- Aggregation
  - the science of **throwing away information**
  - why: *summarize, describe, compress*
- Learning
  - using data to **update our beliefs**
  - why: *test hypotheses, make inferences*
- Sampling
  - **generalizing** our inferences beyond the data at-hand
  - why: *our qs are almost always about more than we can measure*
- Uncertainty
  - characterizing the **limits** of our inferences
  - why: *the world is probabilistic and we can't measure everything!*

Fundamental concepts

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 2087–2097: Review
https://doi.org/10.1080/01621459.2021.1938081

Taylor & Francis
Taylor & Francis Group

Check for updates

What are the Most Important Statistical Ideas of the Past 50 Years?

Andrew Gelman[a] and Aki Vehtari[b]

[a]Department of Statistics, Department of Political Science, Columbia University, New York, NY; [b]Department of Computer Science, Aalto University, Espoo, Finland

ABSTRACT
We review the most important statistical ideas of the past half century, which we categorize as: counterfactual causal inference, bootstrapping and simulation-based inference, overparameterized models and regularization, Bayesian multilevel models, generic computation algorithms, adaptive decision analysis, robust inference, and exploratory data analysis. We discuss key contributions in these subfields, how they relate to modern computing and big data, and how they might be developed and extended in future decades. The goal of this article is to provoke thought and discussion regarding the larger themes of research in statistics and data science.

ARTICLE HISTORY
Received November 2020
Accepted May 2021

KEYWORDS
History of statistics; Data analysis; Statistical computing

Bootstrapping -> *aggregation & sampling*

Simulation -> *sampling & uncertainty*

Multi-level models -> *aggregation & uncertainty*

Algorithms & models -> *aggregation & learning*

Mini-discussion:

*How do these principles apply to a scientific question **you** are interested in?*

- <u>Aggregation</u>
    - the science of **throwing away information**
- <u>Learning</u>
    - using data to **update our beliefs**
- <u>Sampling</u>
    - **generalizing** our inferences beyond the data at-hand
- <u>Uncertainty</u>
    - characterizing the **limits** of our inferences

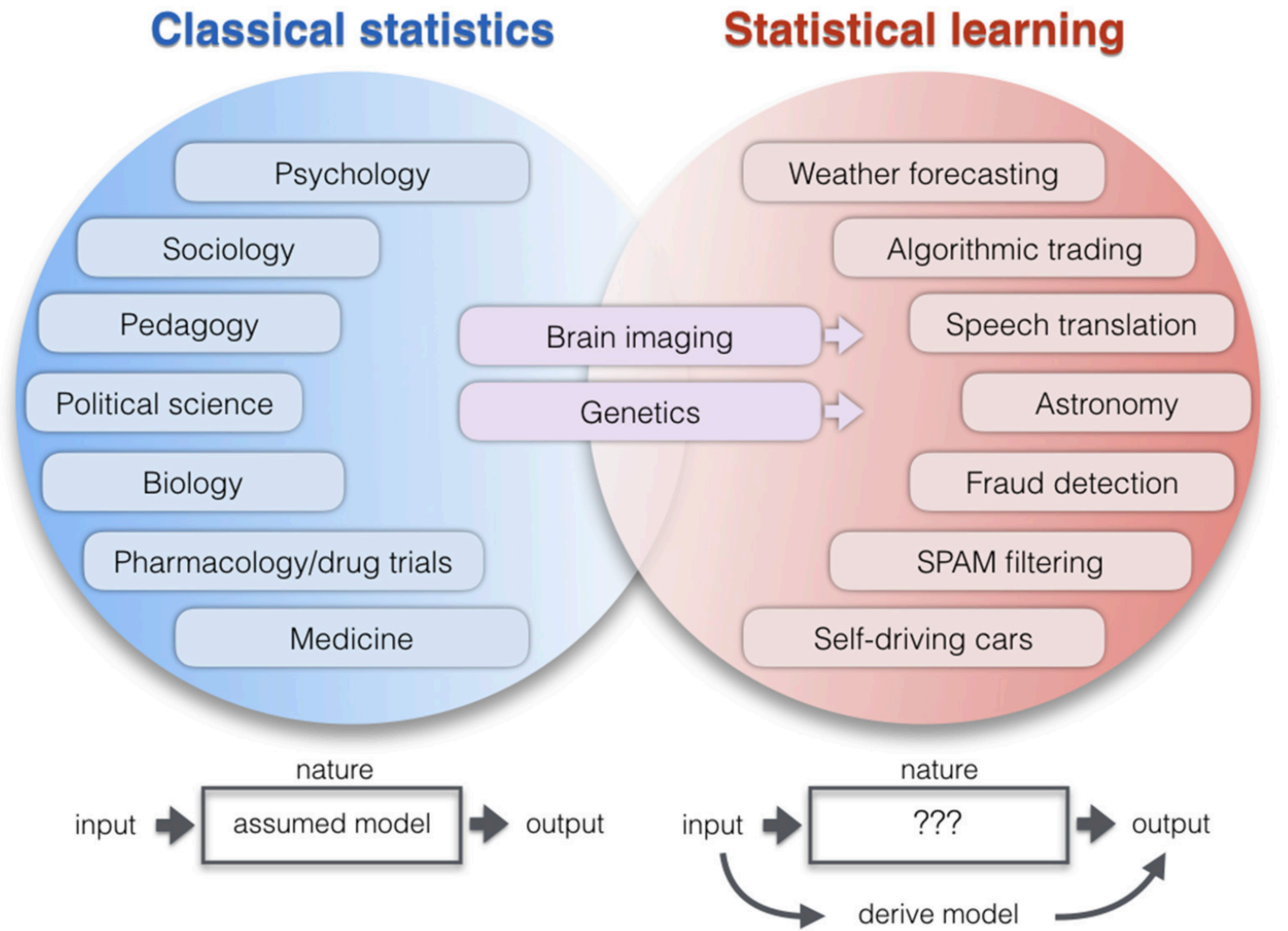
The two cultures of statistics

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

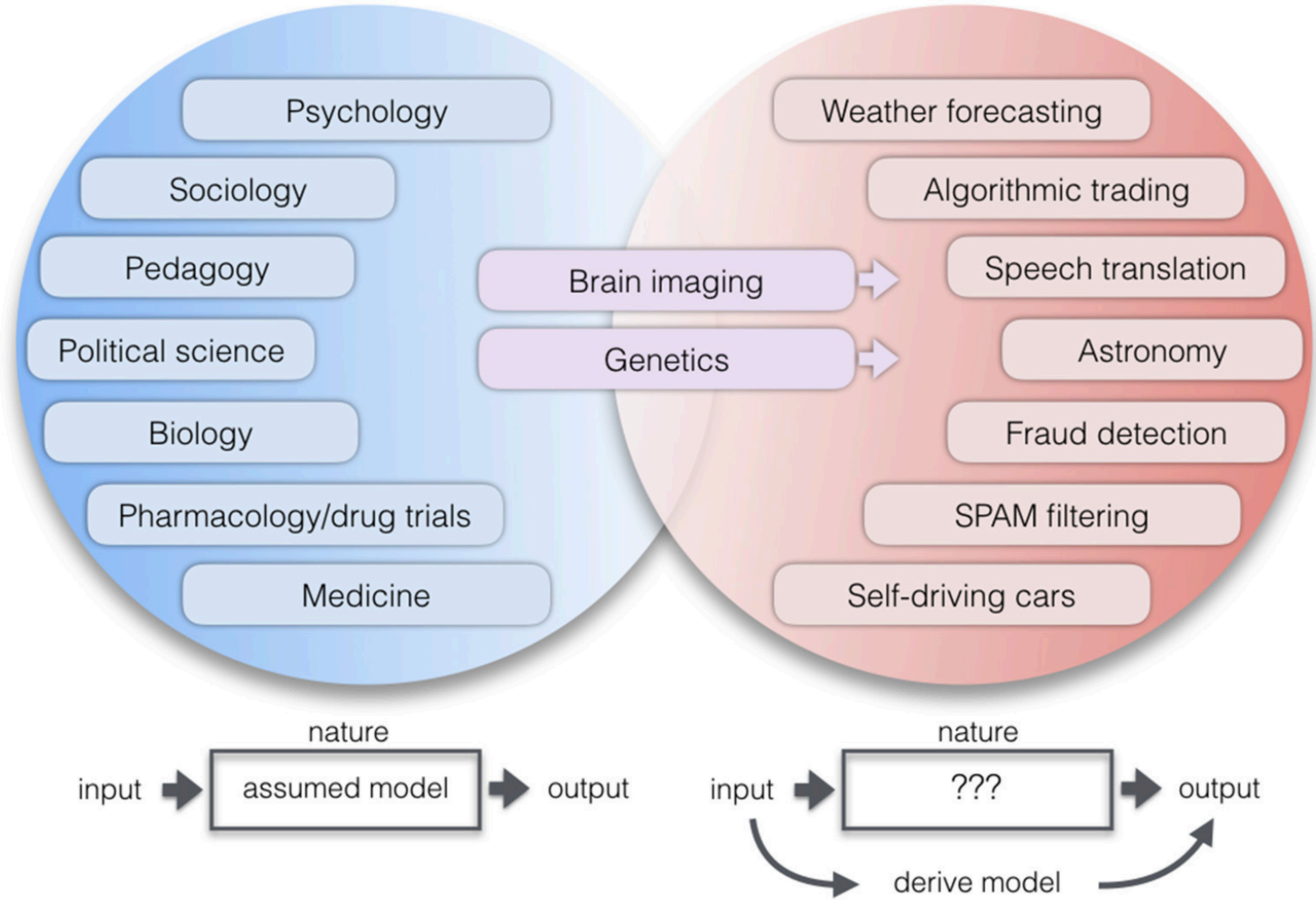
Leo Breiman

Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.



Classical statistics
Statistical learning
Psychology
Sociology
Pedagogy
Political science
Biology
Pharmacology/drug trials
Medicine
Brain imaging
Genetics
Weather forecasting
Algorithmic trading
Speech translation
Astronomy
Fraud detection
SPAM filtering
Self-driving cars
nature
input
assumed model
output
nature
input
???
output
derive model

Explanation　Prediction



Explanation ⟷ **Prediction**

Classical inference

Theory		
null-hypothesis testing & multiple comparisons	bias-variance decomposition	Vapnik-Chervonenkis dimensions & curse of dimensionality
degrees of freedom df		hypothesis space $\mathcal{H}$
asymptotic consistency		finite-sample theorems
Invalidations of inferential process		
double dipping/circular analysis	post-selection inference	data snooping/peeking
Outcome metrics		
(in-sample) p values sensitivity/specificity effect size/power	explained variance metrics AUC/ROC curve confidence intervals	out-of-sample prediction accuracy/precision/recall/F1 scores learning curves certainty estimates via bootstrap
Representative methods		
Student's *t*-test *F*-test ANOVA Binomial test X^2-test linear regression	$\boxed{\text{general(ized) linear model}}$	support vector machines LASSO/ridge regression/elastic net logistic regression nearest neighbors random forests kernel methods („deep“) neural networks

Out-of-sample generalization

Explanation ⟷ **Prediction**

Knowledge guided: pre-specified, simple but inflexible

Pattern guided: “late-commitment” , complex but flexible

Late commitment: Using theoretical assumptions to constrain analysis, not design

In the first step of the empirical cycle, we strive to minimize the theoretical assumptions built into the experimental design. This approach is motivated by the observation that designs, e.g., of fMRI experiments, can be made much more versatile (allowing us to address more neuroscientific questions) at moderate costs in terms of statistical efficiency (for addressing a given question). A general design that can address a 100 questions appears more useful than a restricted design that addresses a single question with slightly greater efficiency.

Statistical power is afforded by combining the evidence – usually by averaging. When we decide on a grouping of experimental events (e.g., for a block design), we commit to a particular way of combining the evidence and thus give up versatility. Ungrouped-events designs allow us to combine the evidence in many different ways *during analysis*. First, this approach allows for exploratory analyses, which can (1) test basic assumptions of a field, (2) usefully direct our attention to larger phenomena (in terms of explained variance), and (3) lead to unexpected discoveries. Second, ungrouped-events designs allow a broad set of theoretically constrained analyses to be performed on the same data. And third, as a consequence, such designs allow us to combine data across studies and research groups in order to address a particular question with a power otherwise unattainable. In the Appendix, we assess this third point, the potential of data sharing within subfields of neuroscientific inquiry, in detail.

Explanation ⟷ **Prediction**

Knowledge guided: pre-specified, simple but inflexible

Pattern guided: “late-commitment” , complex but flexible

Explainable narrative centered on inferences about single *parameters* “beta-hat” problems

Opaque “black box” centered on inferences about out-of-sample *prediction* “y-hat” problems

IV
(independent variable)
(manipulated)

$$\hat{Y} = a + bX + \varepsilon$$

Predicted DV
(dependent variable)
(measured)

Estimated parameter
(coefficient)
(relationship between Y~X)

IV
(independent variable)
(manipulated)

$$\hat{Y} = a + bX + \varepsilon$$

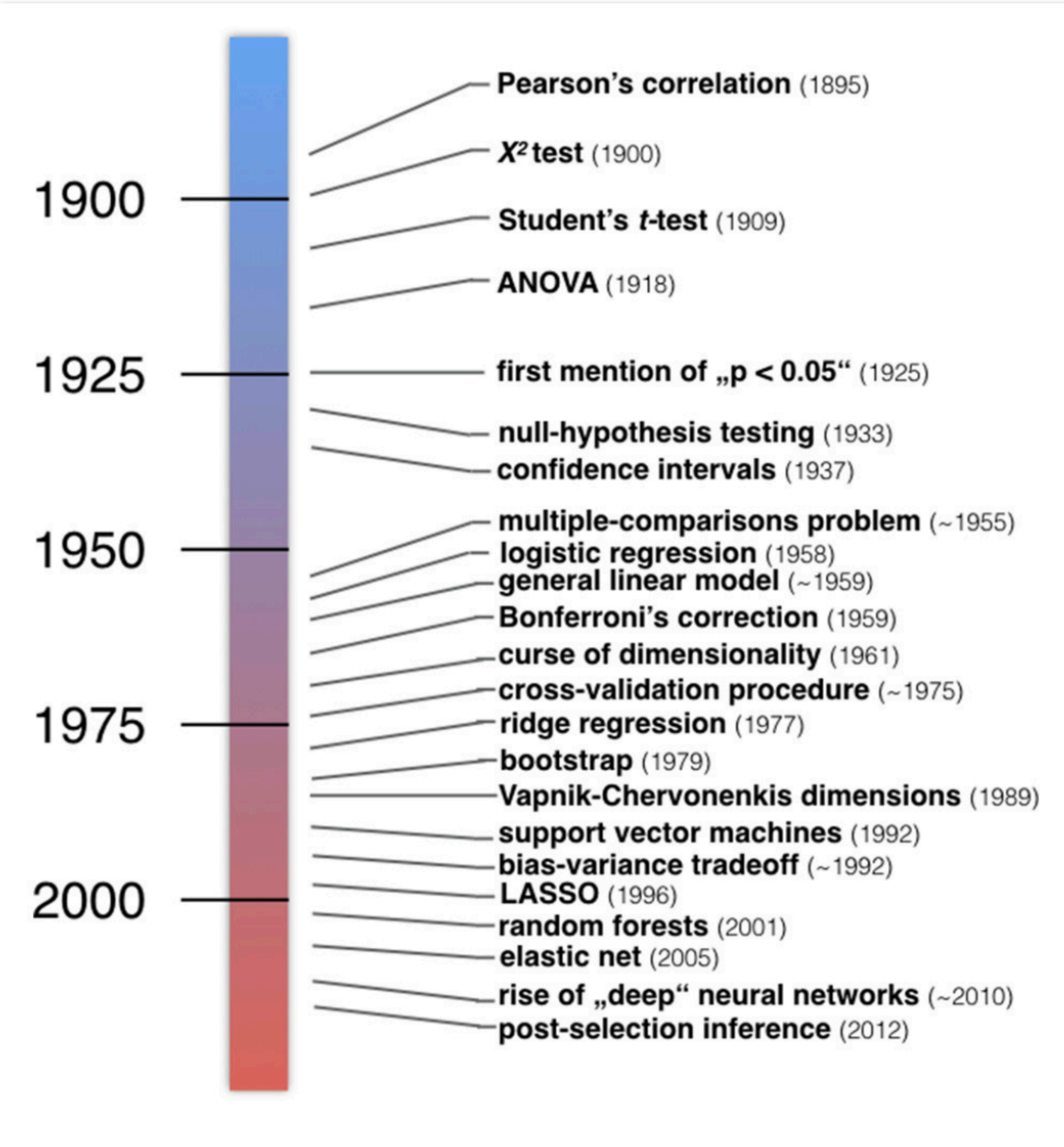
Prediction

Explanation

Explanation ⟷	**Prediction**
Knowledge guided: pre-specified, simple but inflexible	Pattern guided: “late-commitment” , complex but flexible
Explainable narrative centered on inferences about single *parameters* “beta-hat” problems	Opaque “black box” centered on inferences about out-of-sample *prediction* “y-hat” problems
Tools: p-values, ANOVA, t-test, Bayes posterior	Tools: cross-validation, regularization, model comparison
Formally justified: rigorous mathematical and theoretical guarantees, well understood analytic routines and assumptions	Empirically justified: informally validated via resampling and application to unseen data

Explanation	⟷	**Prediction**
„Truth“ is in the model	1	„Truth“ is in the data
„Long data“ (n samples > p variables)	2	„Wide data“ (n samples << p variables)
Make a set of judicious assumptions	3	Make least assumptions possible
Test a model for the data deductively	4	Learn a model from the data inductively
Impose mathematical rigor	5	Let the data speak for themselves
Choose a model before visiting the data	6	Choose a model as the data are visited
More confirmatory than exploratory	7	More exploratory than confirmatory
Model assumptions are largely explicit	8	Model assumptions are largely implicit
Data are often experimental	9	Data are often observational
Tractable models with few parameters	10	Expressive models with many parameters

Why should you care?



But tend to hang out here in practice

We learn up to here

They want/will make us irrelevant

How can we do better?

Perspectives on Psychological Science
1–23

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691617693393
www.psychologicalscience.org/PPS
SAGE

Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Tal Yarkoni and Jacob Westfall
University of Texas at Austin

Abstract
Psychology has historically been concerned, first and foremost, with explaining the causal mechanisms that give rise to behavior. Randomized, tightly controlled experiments are enshrined as the gold standard of psychological research, and there are endless investigations of the various mediating and moderating variables that govern various behaviors. We argue that psychology's near-total focus on explaining the causes of behavior has led much of the field to be populated by research programs that provide intricate theories of psychological mechanism but that have little (or unknown) ability to predict future behaviors with any appreciable accuracy. We propose that principles and techniques from the field of machine learning can help psychology become a more predictive science. We review some of the fundamental concepts and tools of machine learning and point out examples where these concepts have been used to conduct interesting and important psychological research that focuses on predictive research questions. We suggest that an increased focus on prediction, rather than explanation, can ultimately lead us to greater understanding of behavior.

Mini-Discussion

- Can you think about research in your area in terms of prediction vs explanation?
  - Do they pose different questions?
  - Do they require different answers?
- Can you formulate a research question in terms of prediction vs explanation?
- Does one perspective matter more/less to you?

How can we do better?

Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Perspectives on Psychological Science
1–23

DOI: 10.1177/1745691617693393
www.psychologicalscience.org/PPS
SAGE

Tal Yarkoni and Jacob Westfall
University of Texas at Austin

Abstract
Psychology has historically been concerned, first and foremost, with explaining the causal mechanisms that give rise to behavior. Randomized, tightly controlled experiments are enshrined as the gold standard of psychological research, and there are endless investigations of the various mediating and moderating variables that govern various behaviors. We argue that psychology's near-total focus on explaining the causes of behavior has led much of the field to be populated by research programs that provide intricate theories of psychological mechanism but that have little (or unknown) ability to predict future behaviors with any appreciable accuracy. We propose that principles and techniques from the field of machine learning can help psychology become a more predictive science. We review some of the fundamental concepts and tools of machine learning and point out examples where these concepts have been used to conduct interesting and important psychological research that focuses on predictive research questions. We suggest that an increased focus on prediction, rather than explanation, can ultimately lead us to greater understanding of behavior.

How can we do better?

Perspectives on Psychological Science
1–23

DOI: 10.1177/1745691617693393
www.psychologicalscience.org/PPS
SAGE

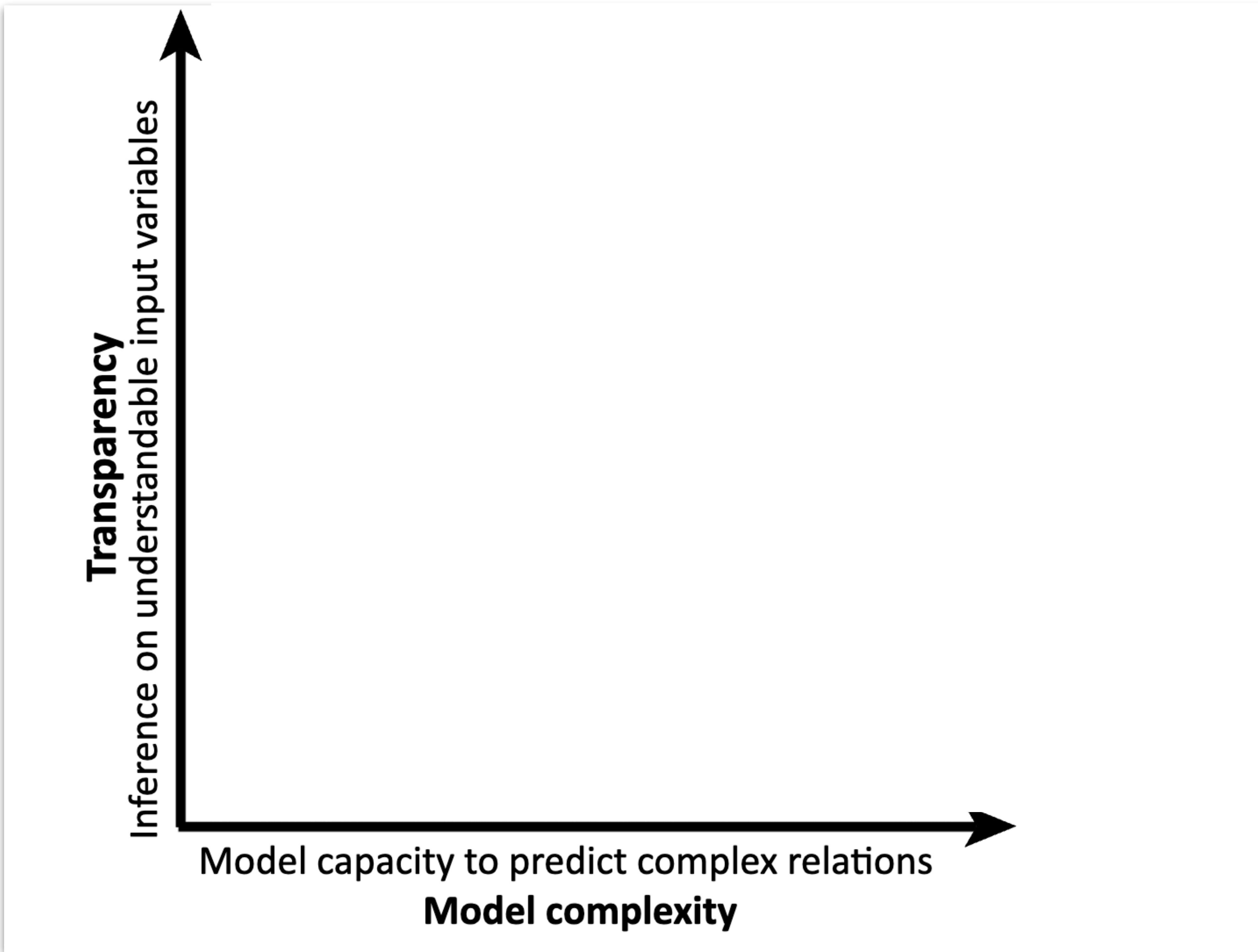
Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Tal Yarkoni and Jacob Westfall
University of Texas at Austin

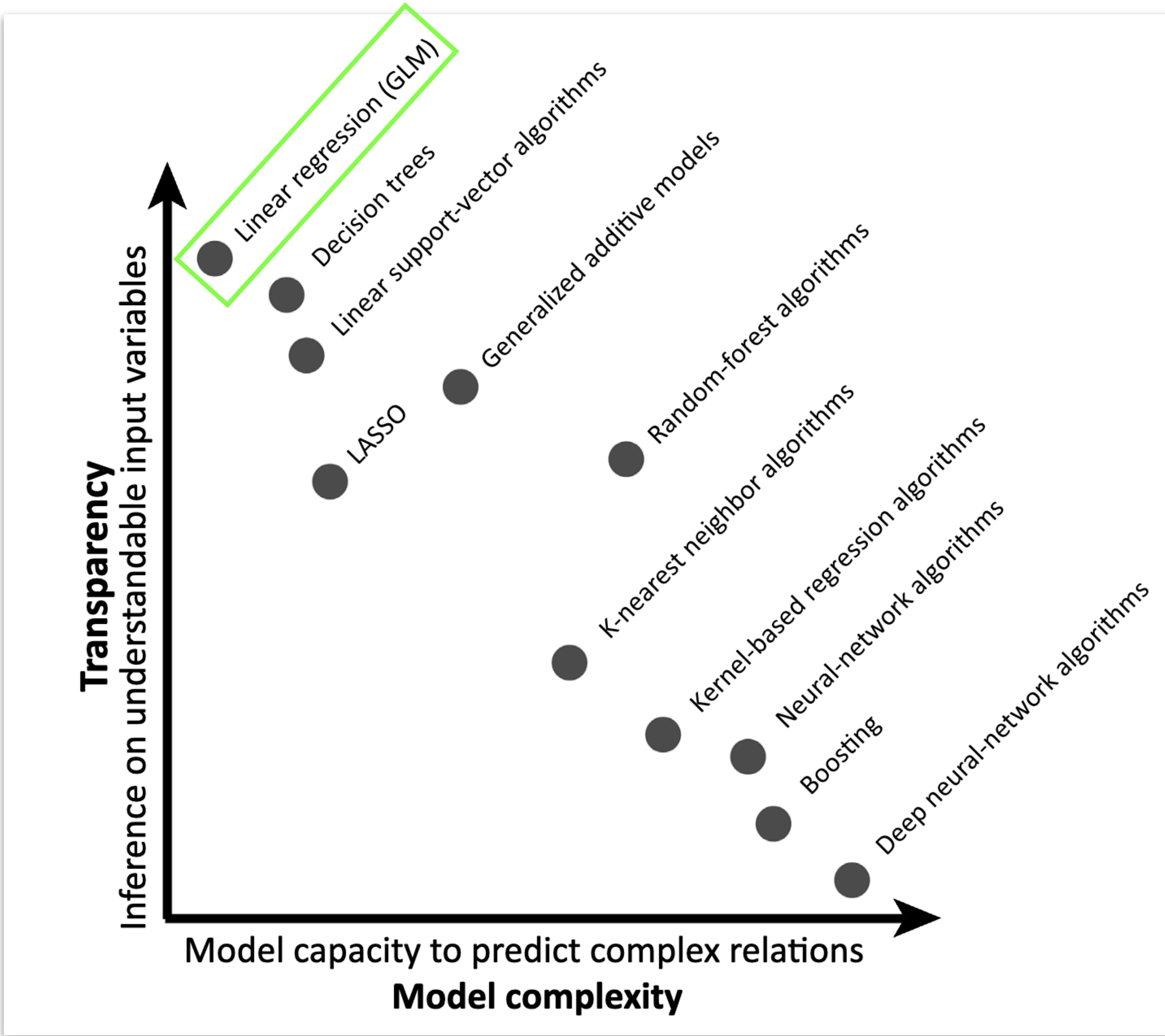
Learn to balance Flexibility & Robustness

Abstract
Psychology has historically been concerned, first and foremost, with explaining the causal mechanisms that give rise to behavior. Randomized, tightly controlled experiments are enshrined as the gold standard of psychological research, and there are endless investigations of the various mediating and moderating variables that govern various behaviors. We argue that psychology's near-total focus on explaining the causes of behavior has led much of the field to be populated by research programs that provide intricate theories of psychological mechanism but that have little (or unknown) ability to predict future behaviors with any appreciable accuracy. We propose that principles and techniques from the field of machine learning can help psychology become a more predictive science. We review some of the fundamental concepts and tools of machine learning and point out examples where these concepts have been used to conduct interesting and important psychological research that focuses on predictive research questions. We suggest that an increased focus on prediction, rather than explanation, can ultimately lead us to greater understanding of behavior.

Complexity vs Understandability



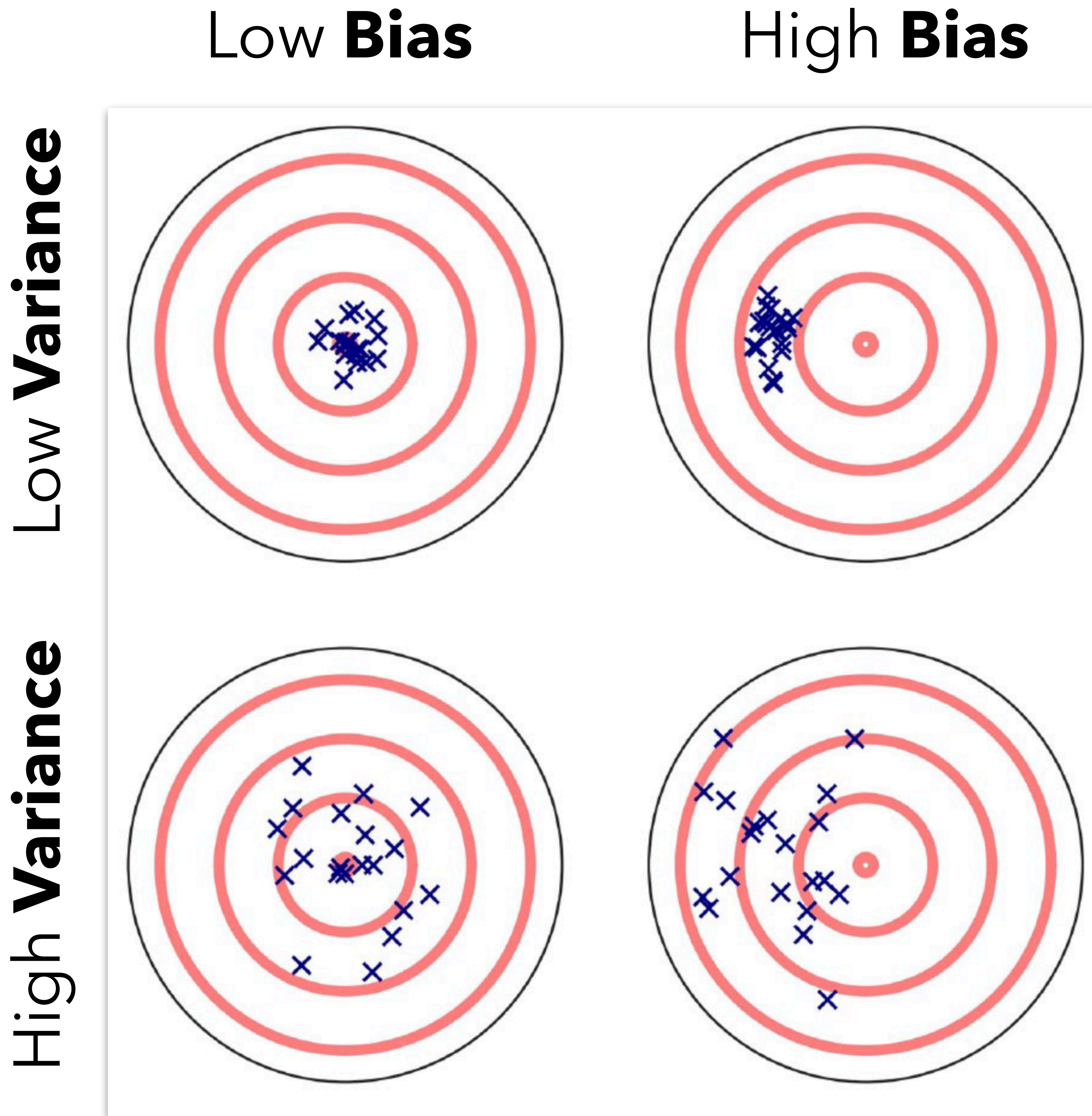
Complexity vs Understandability



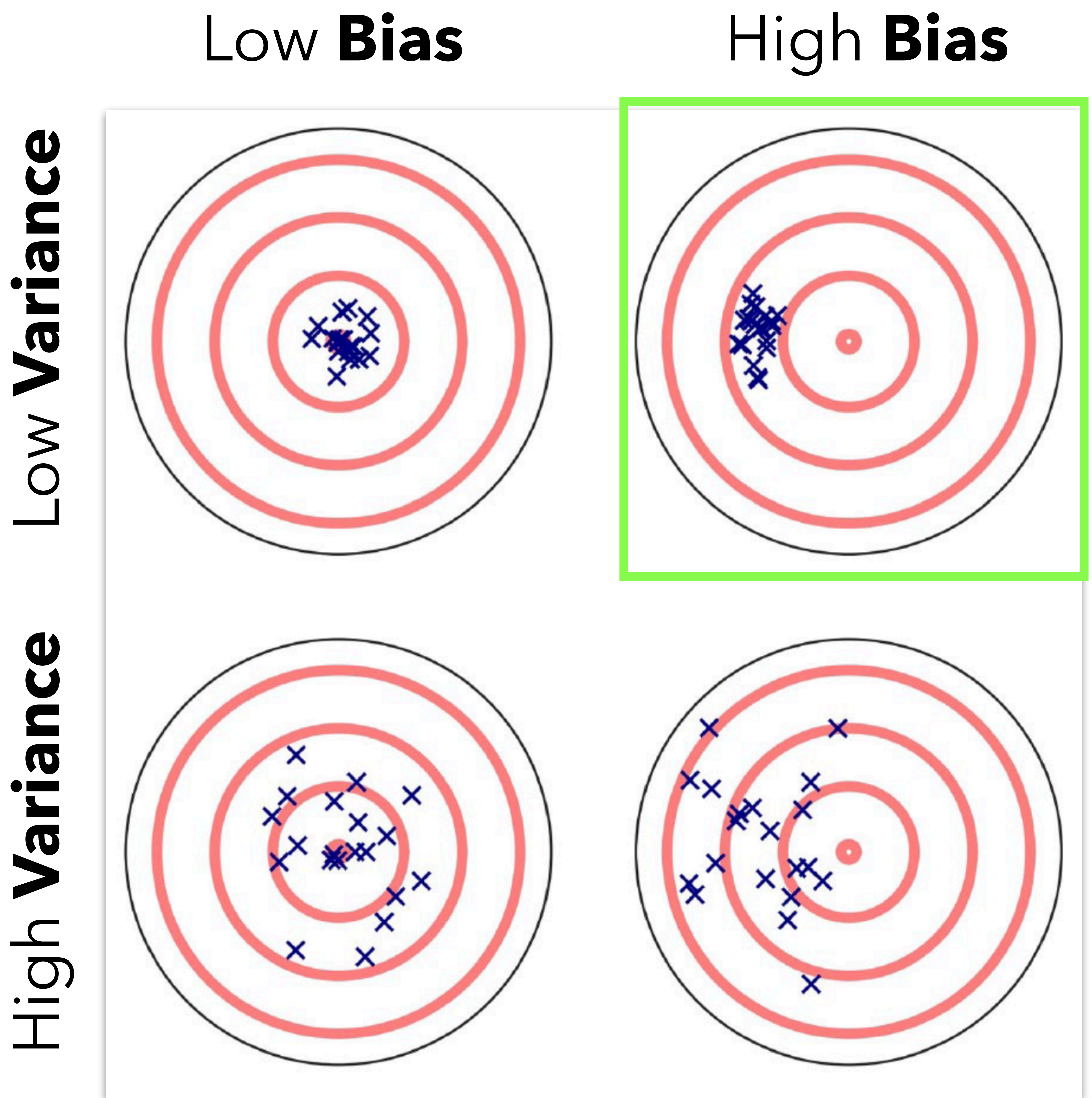
Decomposing error: bias and variance

- In psychology and psychometrics *bias* == error
    - bias is "bad" and we have to "avoid" or "correct" for
- **Bias** is actually only one *type* of error!
    - how "*systematically off*" are we?
- **Variance** quantifies a *different* type of error
    - how much do individual predictions "*bounce around*"?
- **Different models offer different trade-offs!**

Shifting perspective: Bias-Variance trade-off

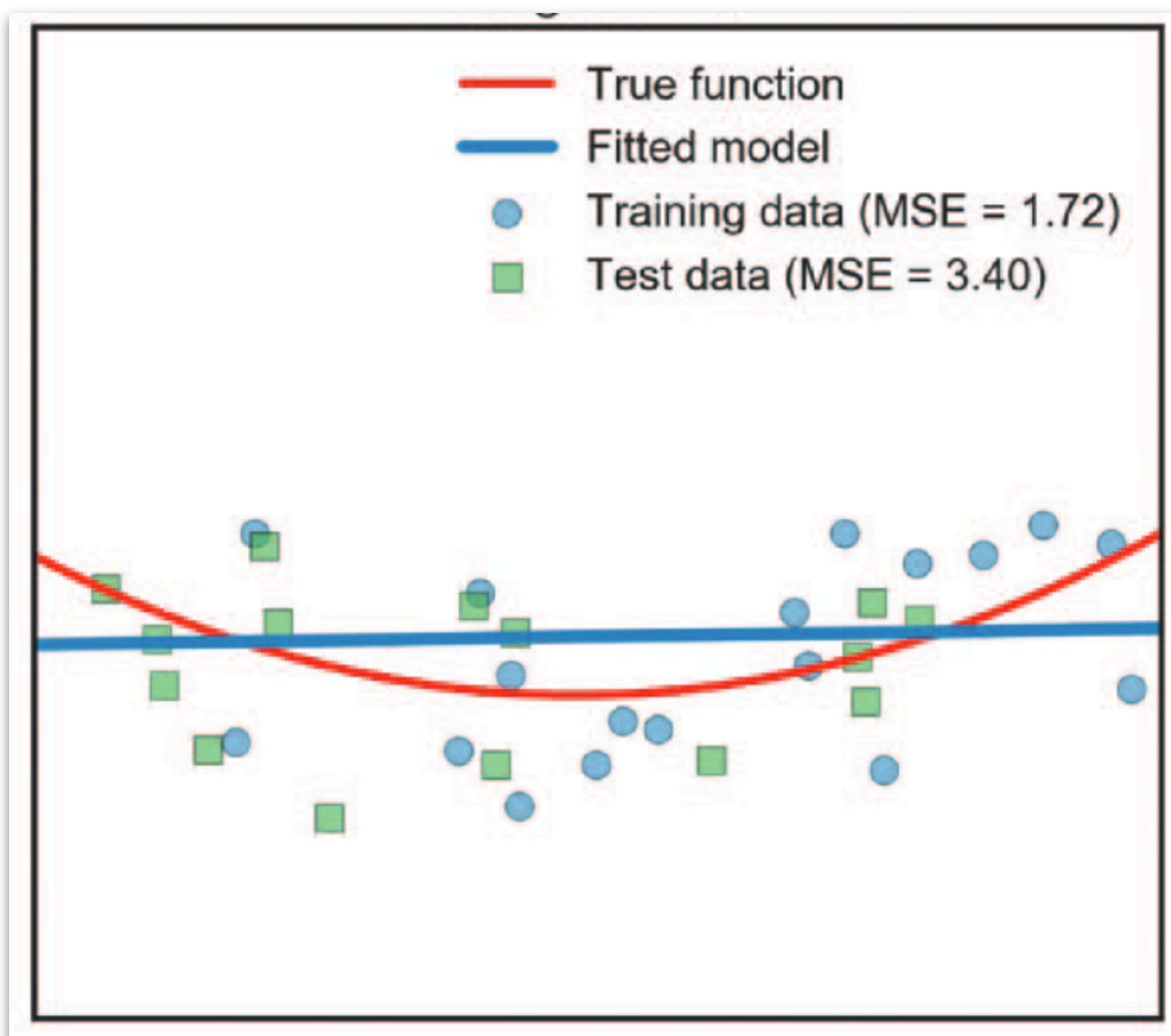


Shifting perspective: Bias-Variance trade-off

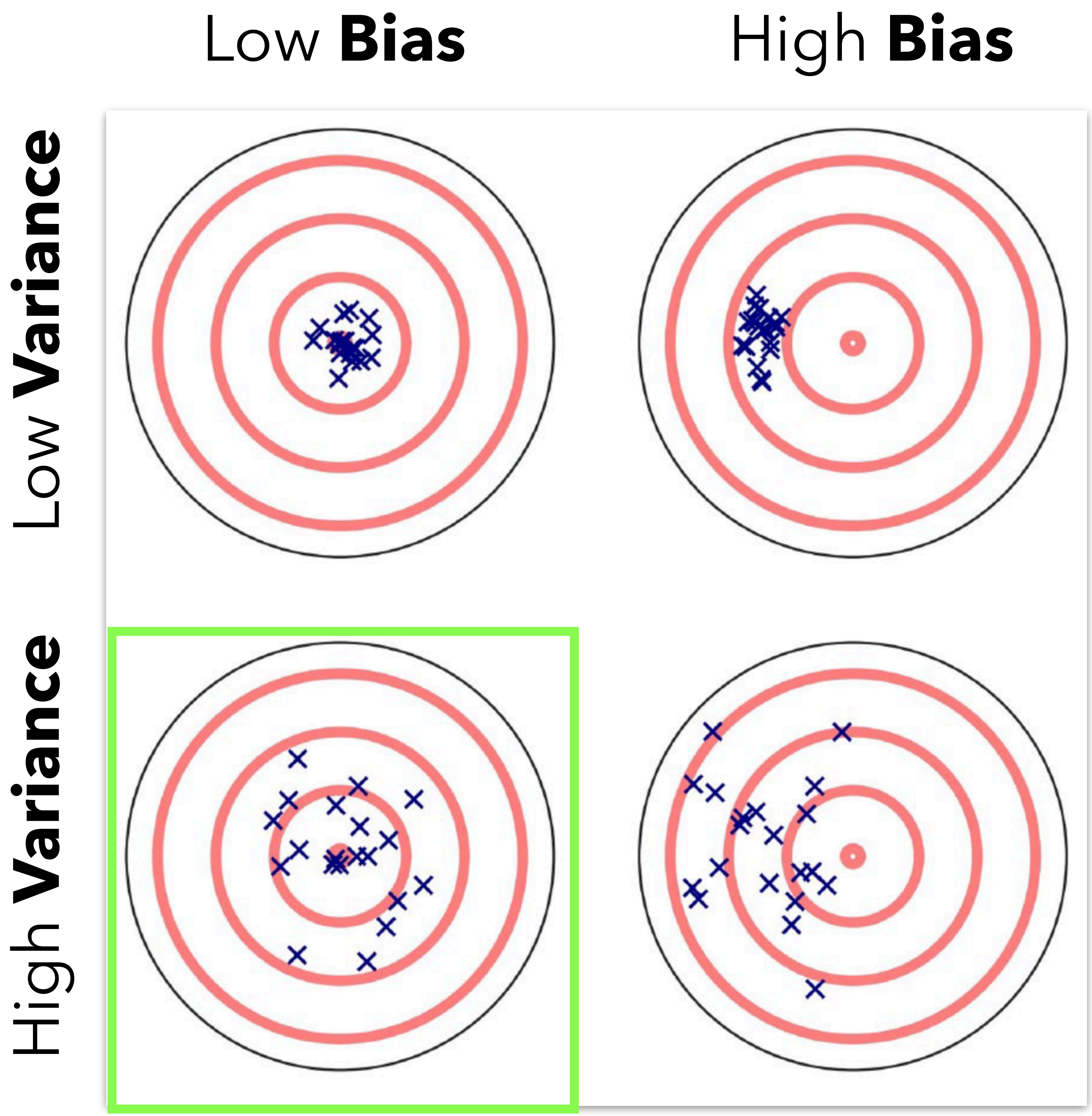


Shifting perspective: Bias-Variance trade-off

- High bias/Low variance
  - "Under-fitting"
  - Being "consistently wrong"
  - Oversimplified but understandable
  - GLM

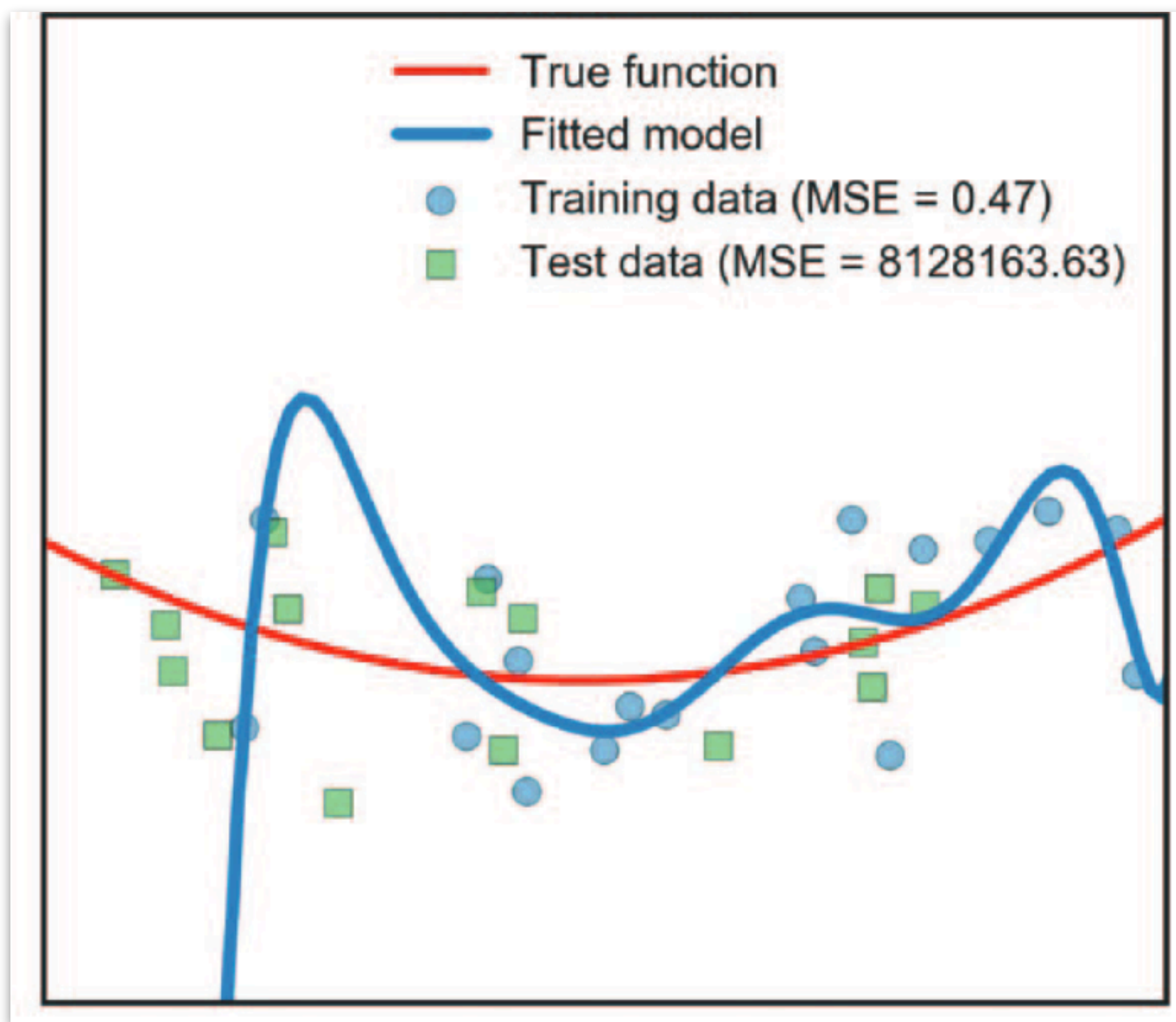


Shifting perspective: Bias-Variance trade-off

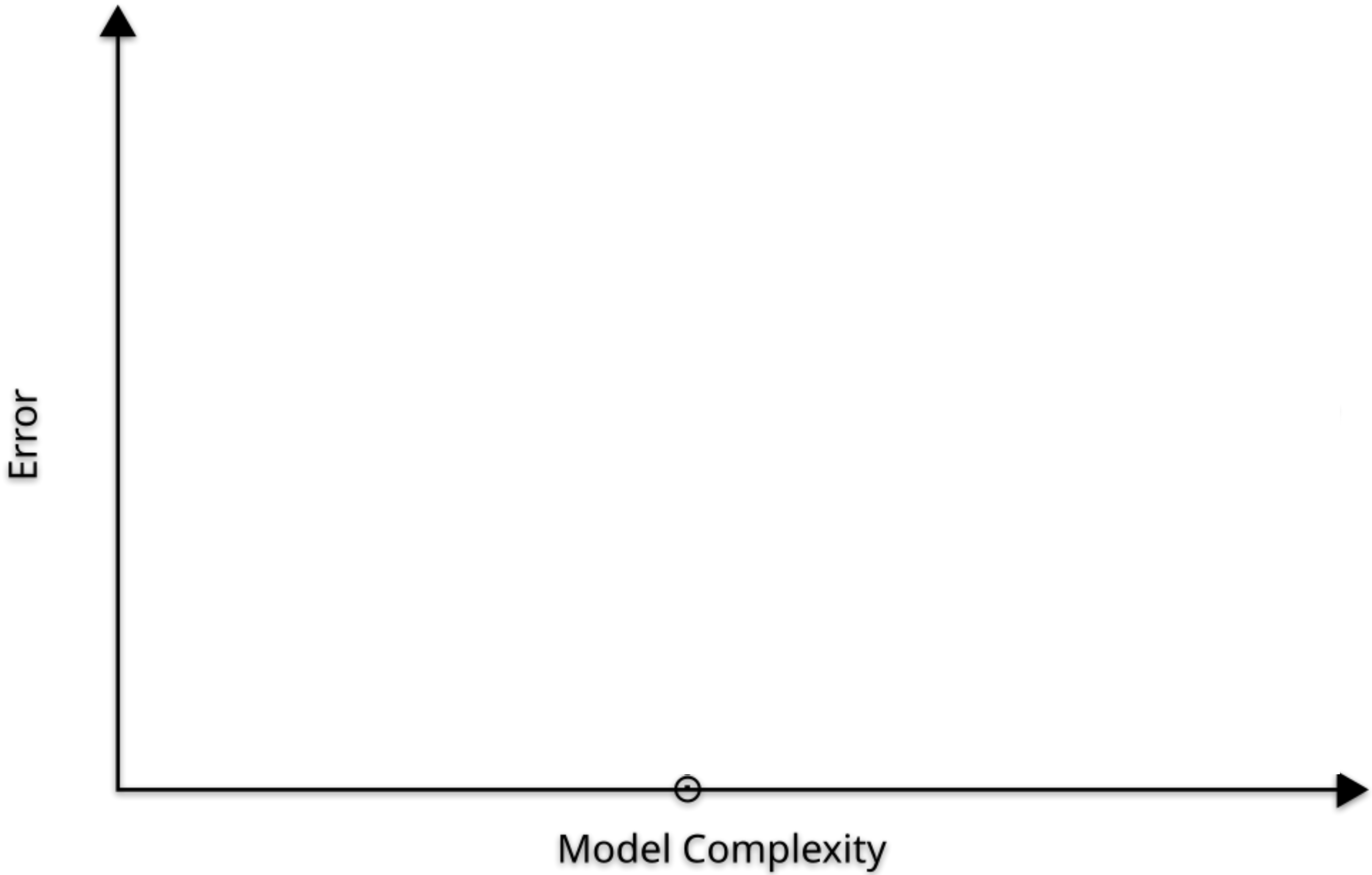


Shifting perspective: Bias-Variance trade-off

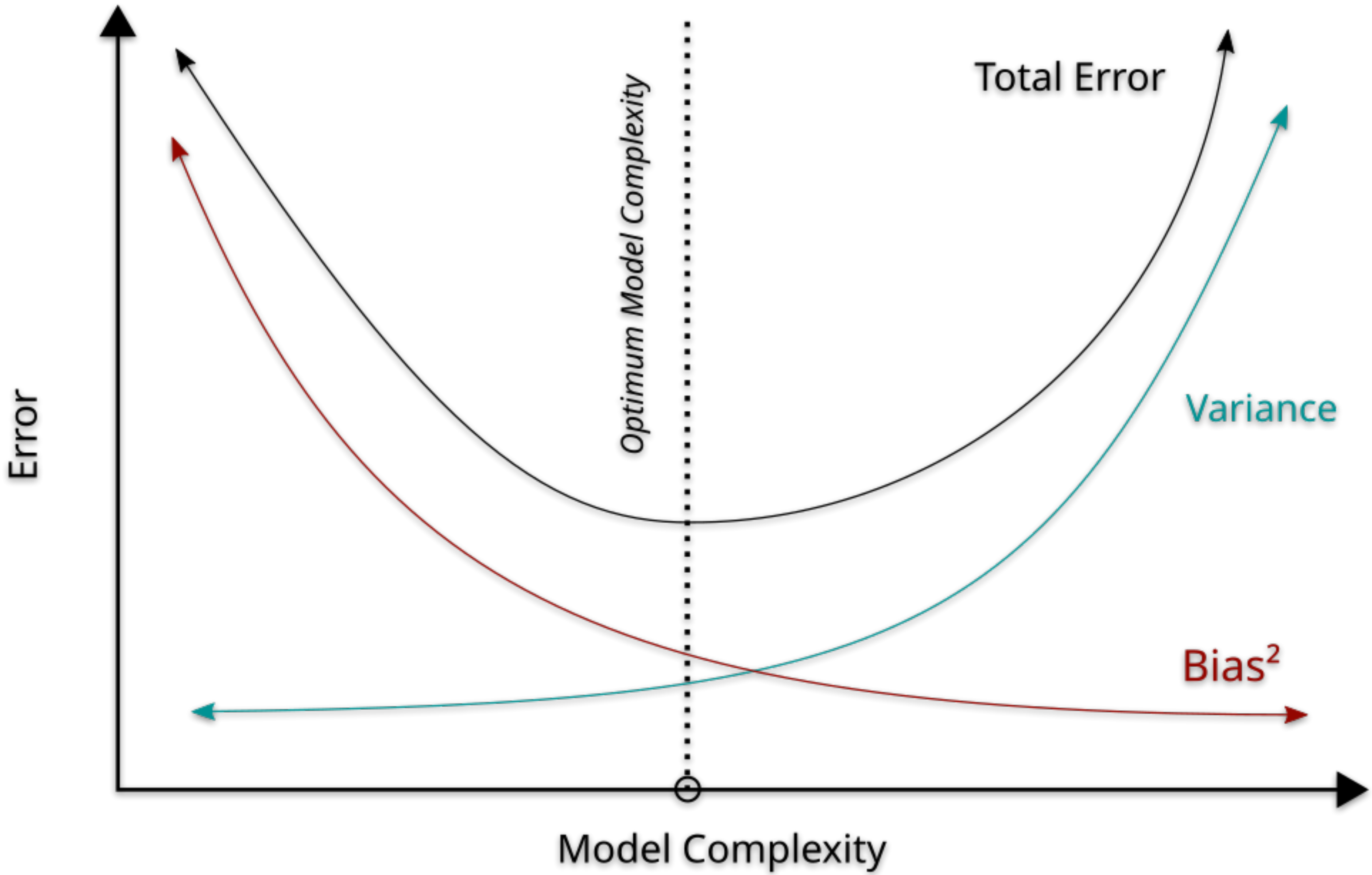
- Low bias/High variance
  - "Over-fitting"
  - Memorizing patterns *and noise* in data
  - Overly complex and unlikely to generalize



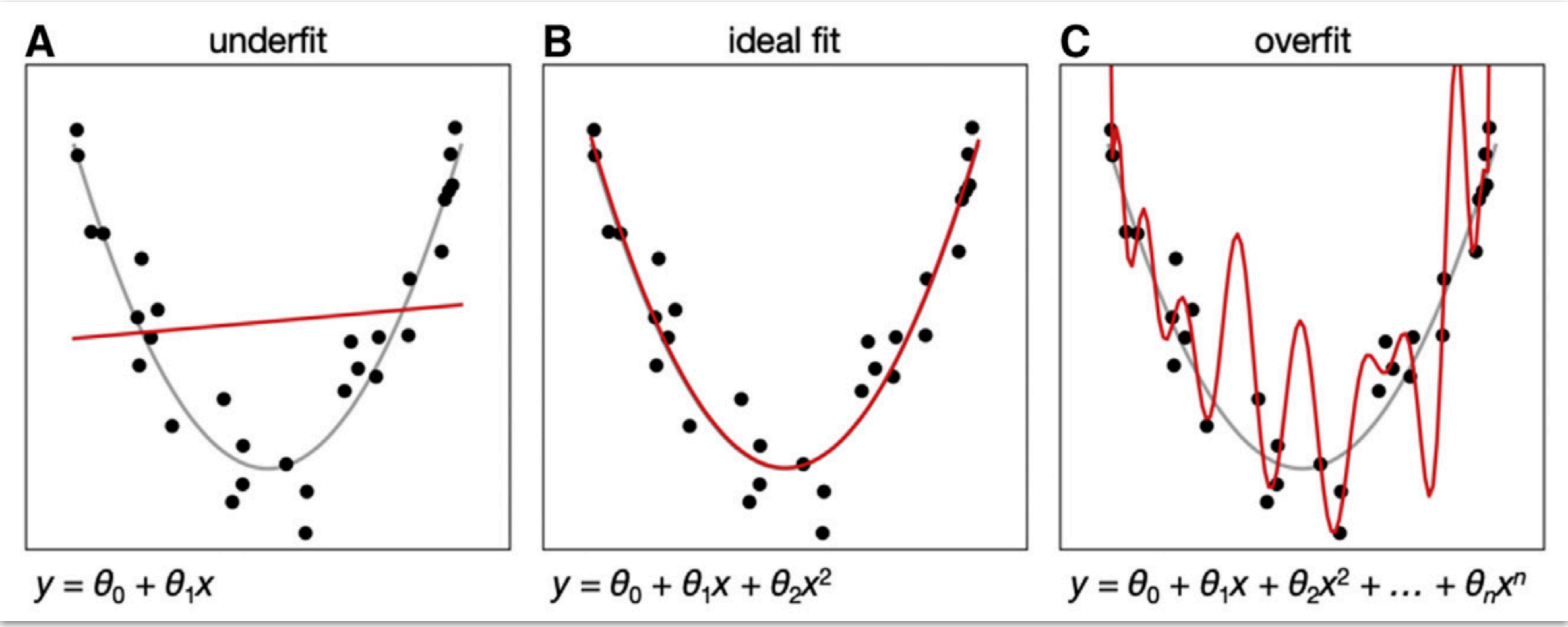
Shifting perspective: Bias-Variance trade-off



Shifting perspective: Bias-Variance trade-off

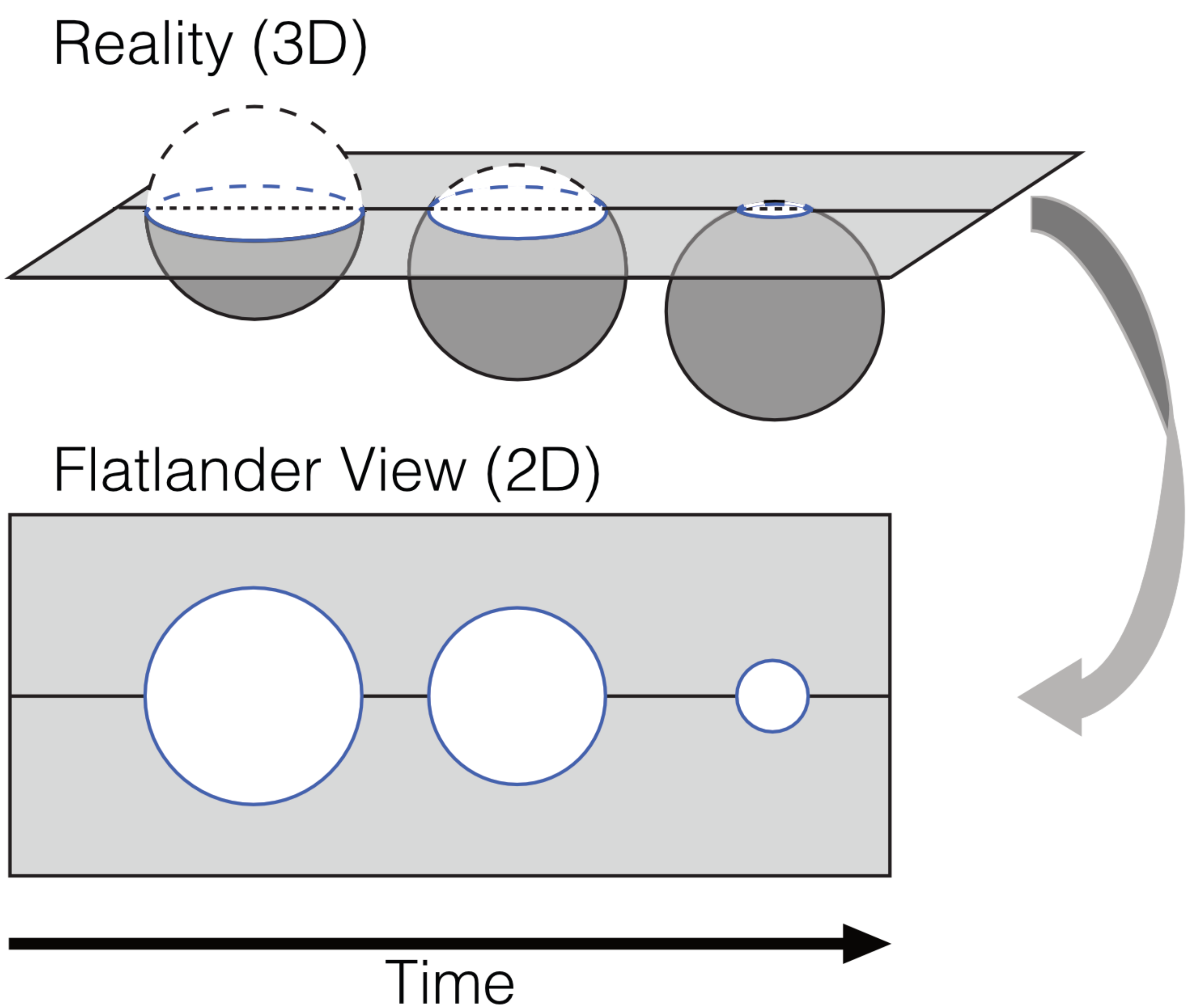


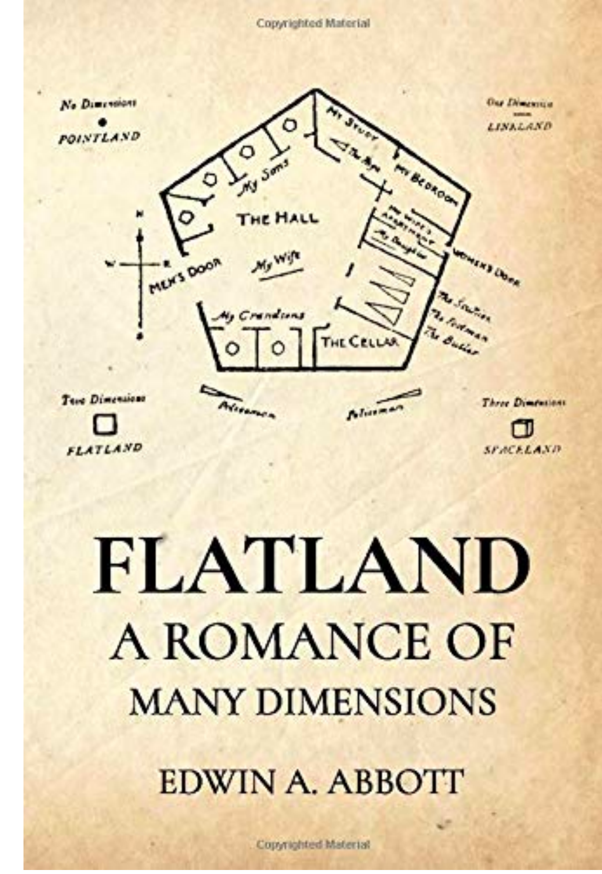
Shifting perspective: Bias-Variance trade-off



Why should you care?

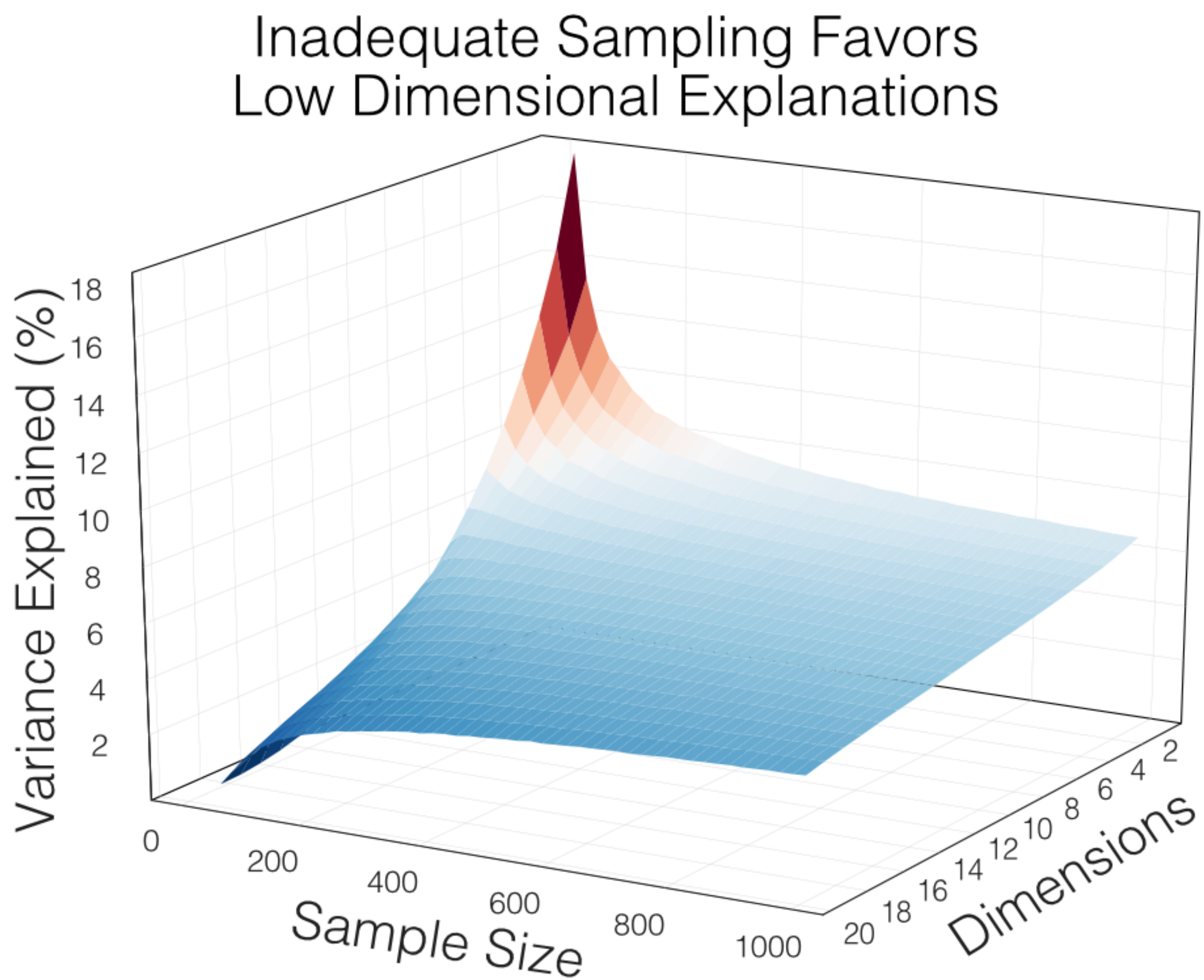
The Flatland Fallacy





Jolly & Chang, 2019

The Flatland Fallacy



Jolly & Chang, 2019

Ecological Validity

"~~Ecological validity~~" — coined by Egon Brunswik in 1947

…to mean something else (¬_¬)

Representative design

Ecological generalizability demands a "representative sampling of situations" where "situational instances in an ecology are analogous to individuals in a population."

Experimental control vs Representative Design

How Close Are We to Understanding V1?

Bruno A. Olshausen
baolshausen@ucdavis.edu
Redwood Neuroscience Institute, Menlo Park, CA 94025, and Center for Neuroscience, University of California at Davis, Davis, CA 95616, U.S.A.

David J. Field
dif3@cornell.edu
Department of Psychology, Cornell University, Ithaca, NY 14853, U.S.A.

"We can rightfully claim to understand only 10% to 20% of how V1 actually operates under normal conditions."

	Biased Sampling	Biased Stimuli	Biased Theories	Interdependence and Context	Ecological Deviance
Problem	Large neurons; visually responsive neurons; neurons with high firing rates	Use of reduced stimuli such as bars, spots, and gratings	Simple/complex cells; data-driven theories	Influence of intracortical input; effect of context; synchrony	Responses to natural scenes deviate from predictions of standard models
Solution	Use chronically implanted electrodes; parallel recording arrays	Use natural scenes, ecologically relevant stimuli	Consider more functional/computational theories that solve problems of vision	Examine how context affects responses in natural scenes	Develop models that can account for responses to natural images

Mini-Discussion: Practical Implications

- How *representative* are experimental designs in your field?
    - Does it matter? Why or why not?
- How do you think Psychology’s understanding of statistics has/has not influence representative design?
- Is Psychology building a robust and cumulative science?
    - Why or why not?
- What kinds of explanations are we “optimizing” for in Psychology?
    - those that we can “understand” and “narrativ-ize”?
    - those that get us closer to a “true” description of the world?

For next time:

1. Please Slack me your Github usernames!
2. Look out for an announcement regarding readings & tutorial notebooks for next week

Thanks!

see you on **Monday @2pm**

make sure to bring your laptop!