Bayesian Modeling Hub Demo

Jared Murray, January 2025

Our Agenda

Introduce new software for fitting Bayesian tree models (stochtree)

Very brief overview Bayesian tree models

Specializing these to causal inference

Demo!

Stochtree

https://stochtree.ai

StochTree

Search

StochasticTree/stochtree
19 6

Home | Getting Started | R Package | Python Package | C++ Core API and Architecture

- Getting Started
  - Python Package
    - Quick start
    - Virtual environment installation
  - R Package
  - C++ Core
    - Compilation
    - Xcode

Getting Started

`stochtree` is composed of a C++ "core" and R / Python interfaces to that core. Details on installation and use are available below:

Python Package

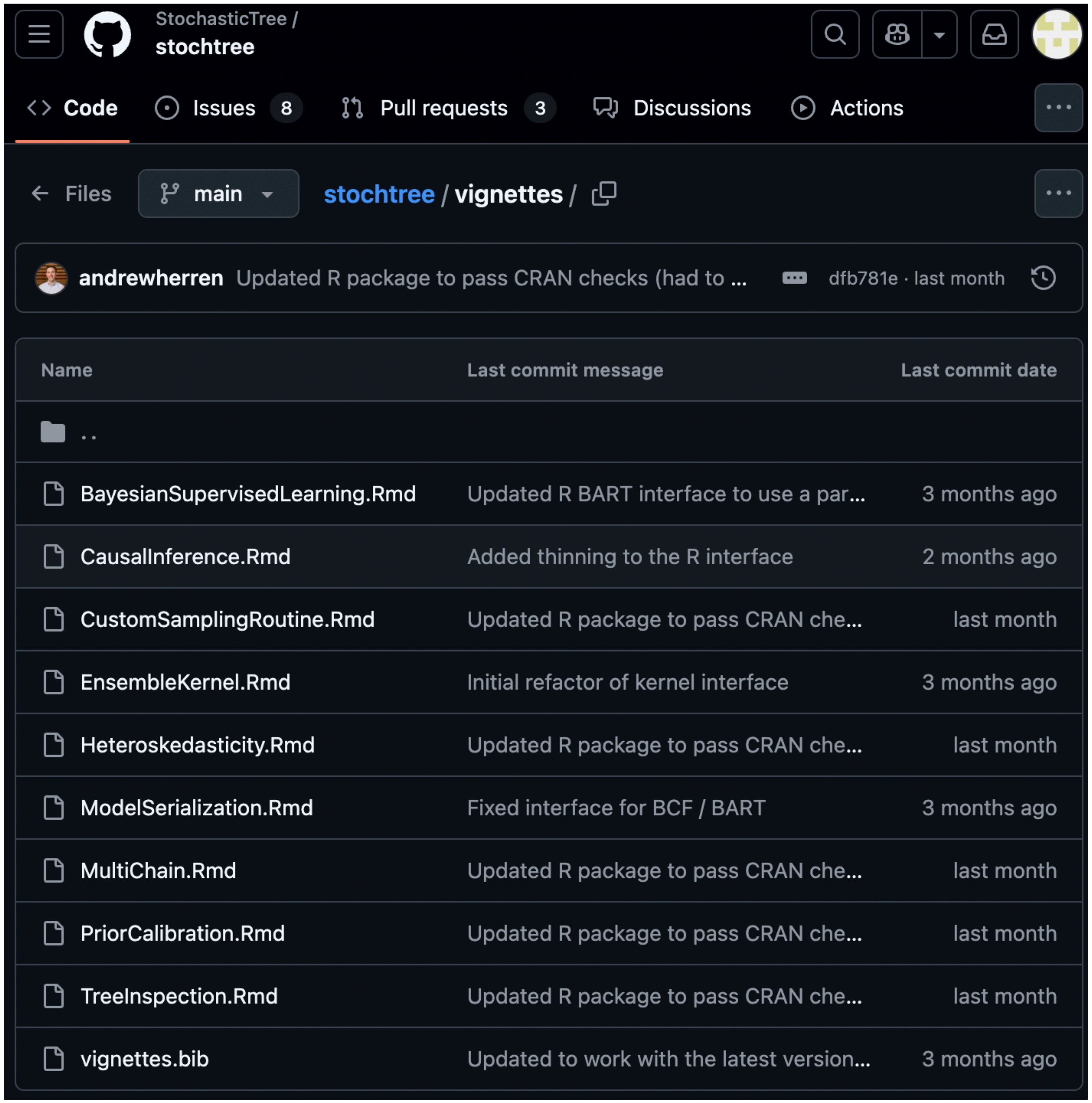
The python package is not yet on PyPI but can be installed from source using pip's git interface. To proceed, you will need a working version of git and python 3.8 or greater (available from several sources, one of the most straightforward being the anaconda suite).

Quick start

Without worrying about virtual environments (detailed further below), `stochtree` can be installed from the command line

```
pip install numpy scipy pytest pandas scikit-learn pybind11
pip install git+https://github.com/StochasticTree/stochtree.git
```

Vignettes for advanced features



Core functionality

Embedding BART in custom models

Heteroskedastic errors

Combining multiple MCMC chains

Docs for Complete APIs

https://stochtree.ai/

StochTree R API Reference

Overview of the `stochtree` R library's key classes and functions, built as a self-contained doc site using the pkgdown format. The `stochtree` interface is divided into two "levels":

1. "High level": end-to-end implementations of stochastic tree ensembles for supervised learning (BART / XBART) and causal inference (BCF / XBCF).
   a. The BART (supervised learning) interface is documented here.
   b. The BCF (causal inference) interface is documented here.
2. "Low level": we provide access to most of the C++ sampling objects and functionality via R, which allow for custom sampling algorithms and integration of other model terms. This interface consists broadly of the following components:
   a. Data API: loading and storing in-memory data needed to train `stochtree` models.
   b. Forest API: creating, storing, modifying, and sampling ensembles of decision trees that underlie all `stochtree` models.
   c. Serialization API: serializing models to JSON (files or in-memory strings).
   d. Random Effects API: sampling from additive random effects models.

StochTree Python API Reference

Overview of the `stochtree` python library's key classes and functions.

The `stochtree` interface is divided into two "levels":

1. "High level": end-to-end implementations of stochastic tree ensembles for supervised learning (BART / XBART) and causal inference (BCF / XBCF). Both interfaces are designed to mirror the scikit-learn estimator style, with the `.fit()` method replaced by a `.sample()` method.
   a. The BART (supervised learning) interface is documented here.
   b. The BCF (causal inference) interface is documented here.
2. "Low level": we provide access to most of the C++ sampling objects and functionality via Python, which allow for custom sampling algorithms and integration of other model terms. This interface is documented here and consists broadly of the following components:
   a. Data API: loading and storing in-memory data needed to train `stochtree` models.
   b. Forest API: creating, storing, and modifying ensembles of decision trees that underlie all `stochtree` models.
   c. Sampler API: sampling from stochastic tree ensemble models as well as several supported parametric models.
   d. Utilities API: seeding a C++ random number generator, preprocessing data, and serializing models to JSON (files or in-memory strings).

The stochtree advantage

Cross-platform (R/Python/C++)

Accelerated ("X") versions of algorithms for large datasets

Extensible, and (relatively) easy to embed in larger models

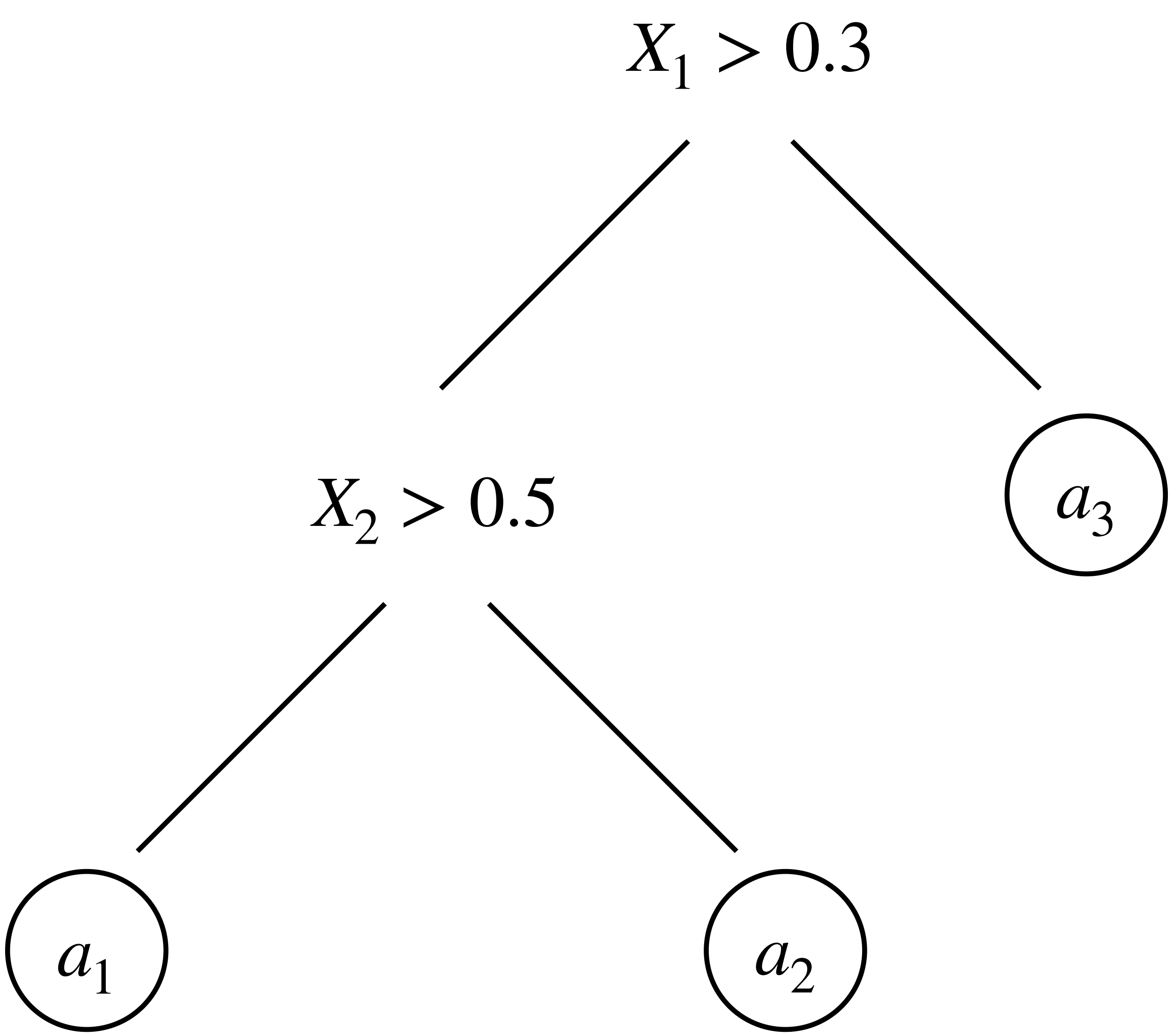
Today: Introduction and demo

Focusing on simplest case (evaluating randomized controlled trial)

Observational studies are a simple extension

Ongoing work: Multiarm treatments, bandits/adaptive trials, multiple outcomes, …

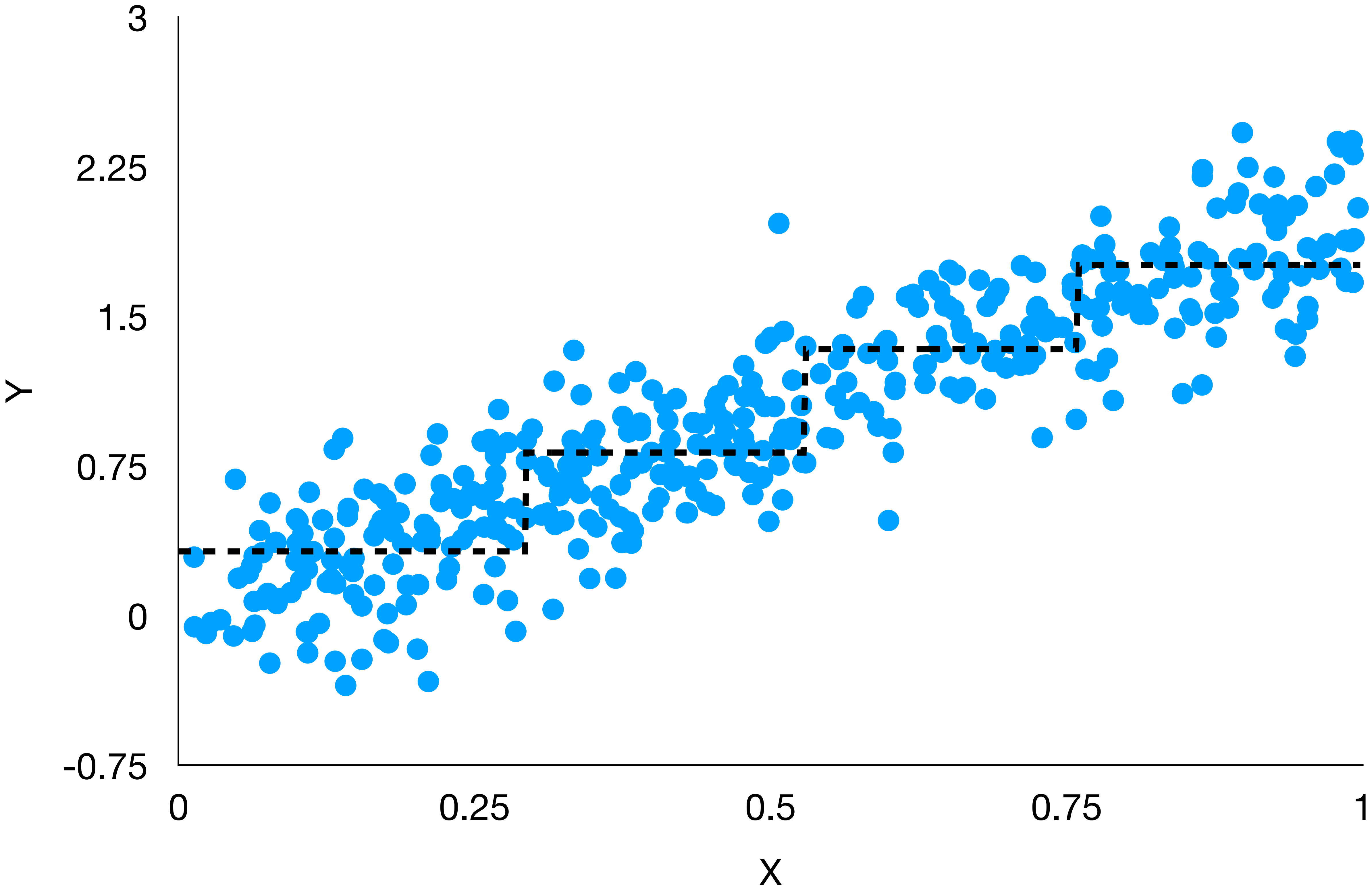
Trees are a simple but powerful machine learning tool …



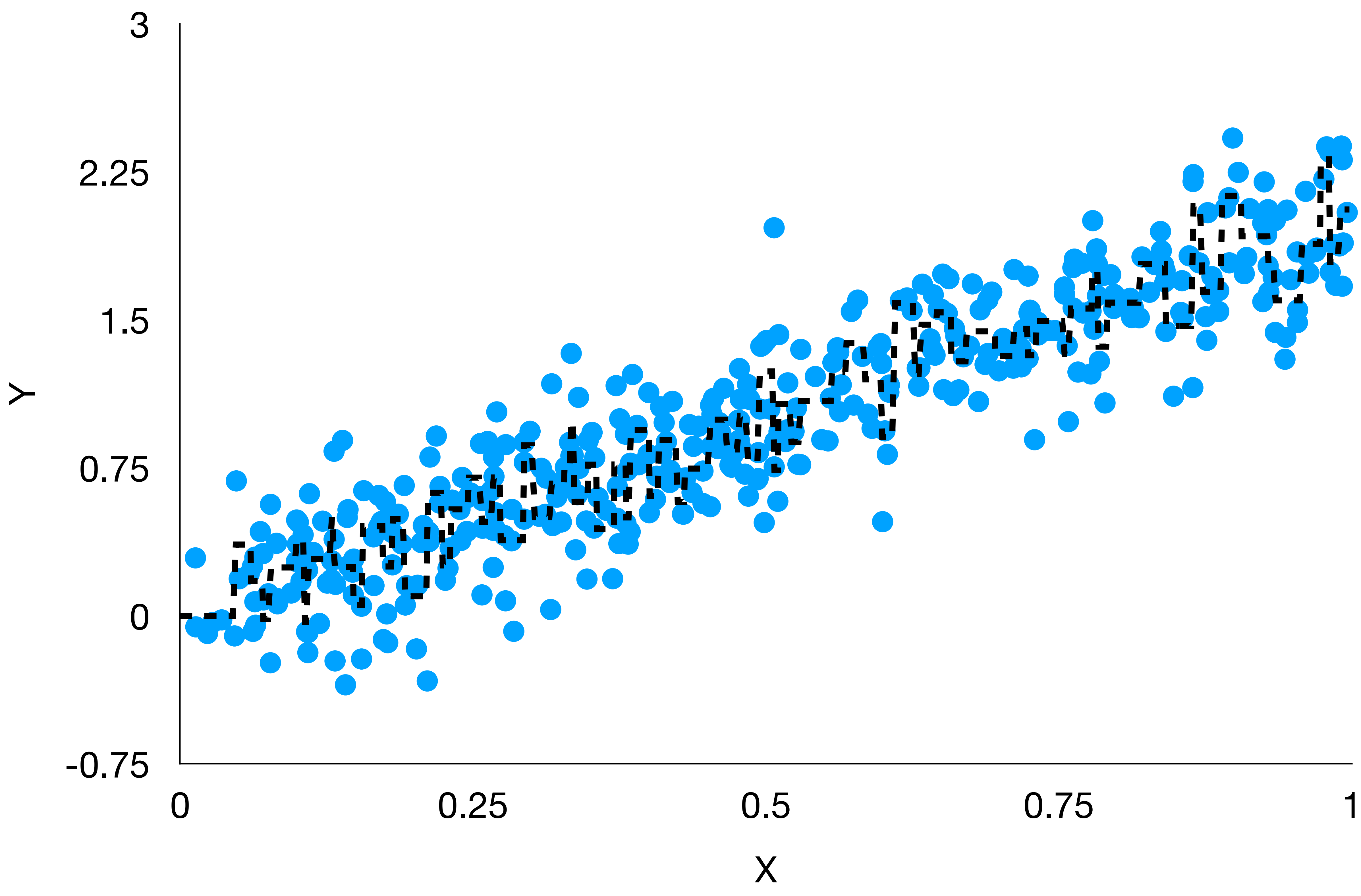
- Algorithm partitions the data - no need to specify model form like regression or neural networks
- Easy to understand model as a series of “if feature 1 < 0.5 then predict a_j” statements

… but they present a difficult overfitting tradeoff

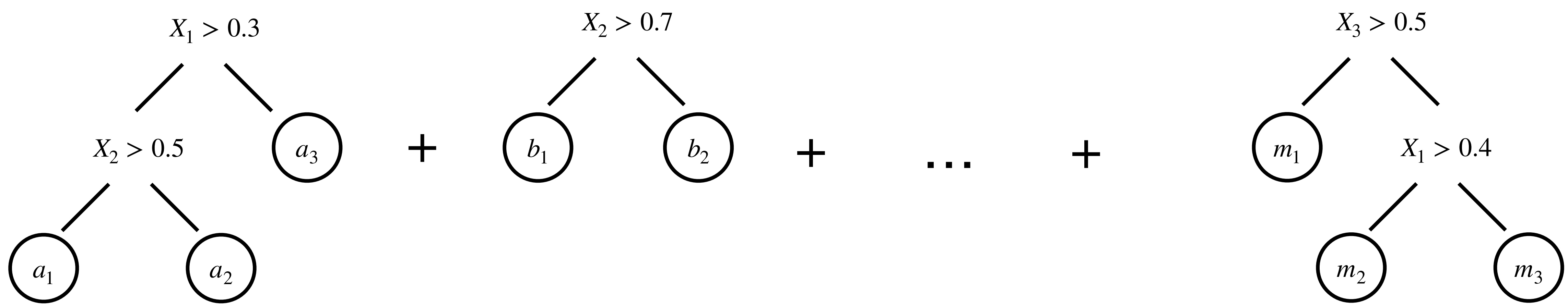
Simple decision trees don't capture much complexity



Complex decision trees unlikely to generalize well!

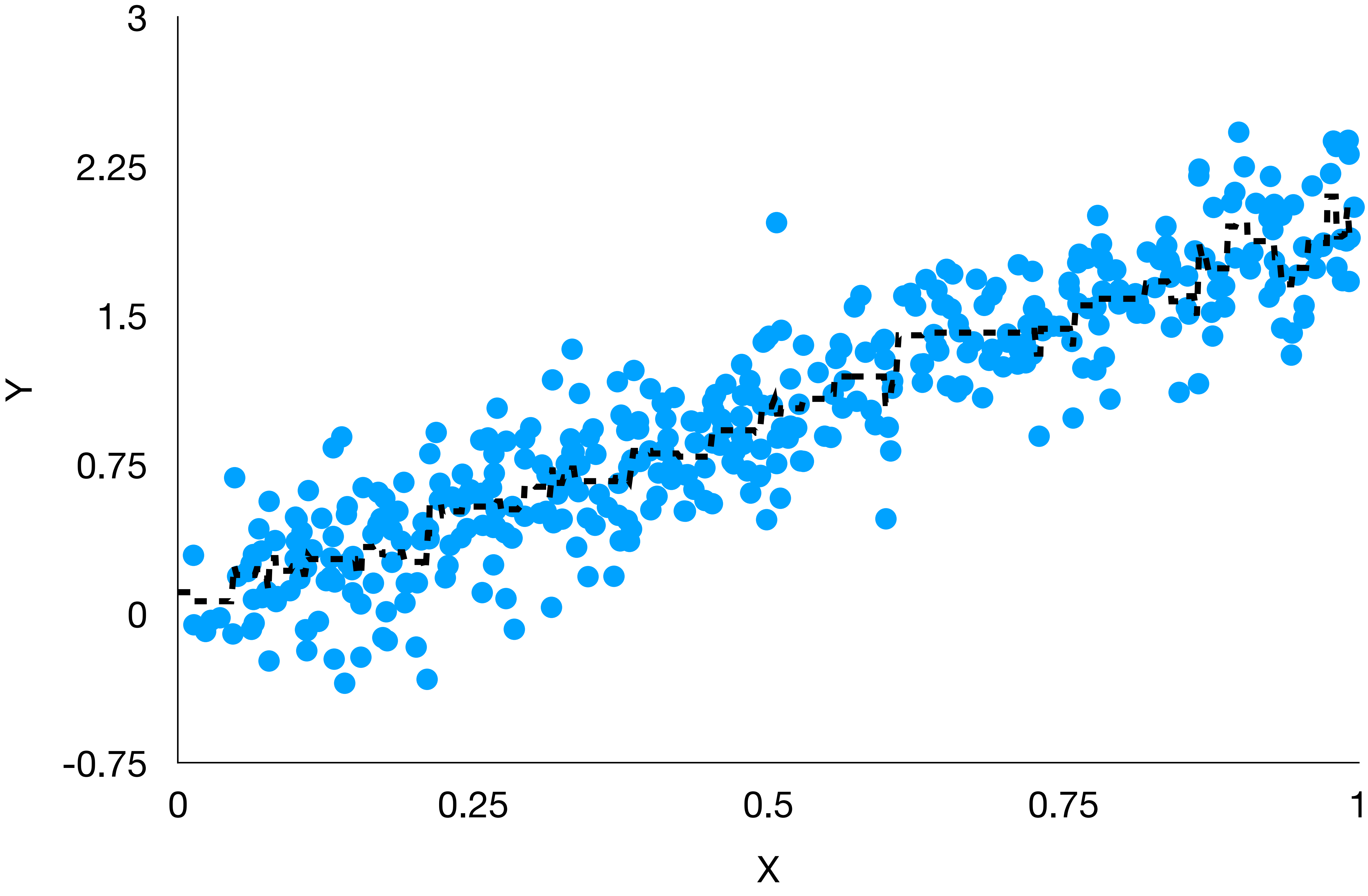


Solution: Ensembling decision trees



This includes common methods like random forest and gradient boosting, implemented in scikit-learn, xgboost, lightgbm, ranger, …

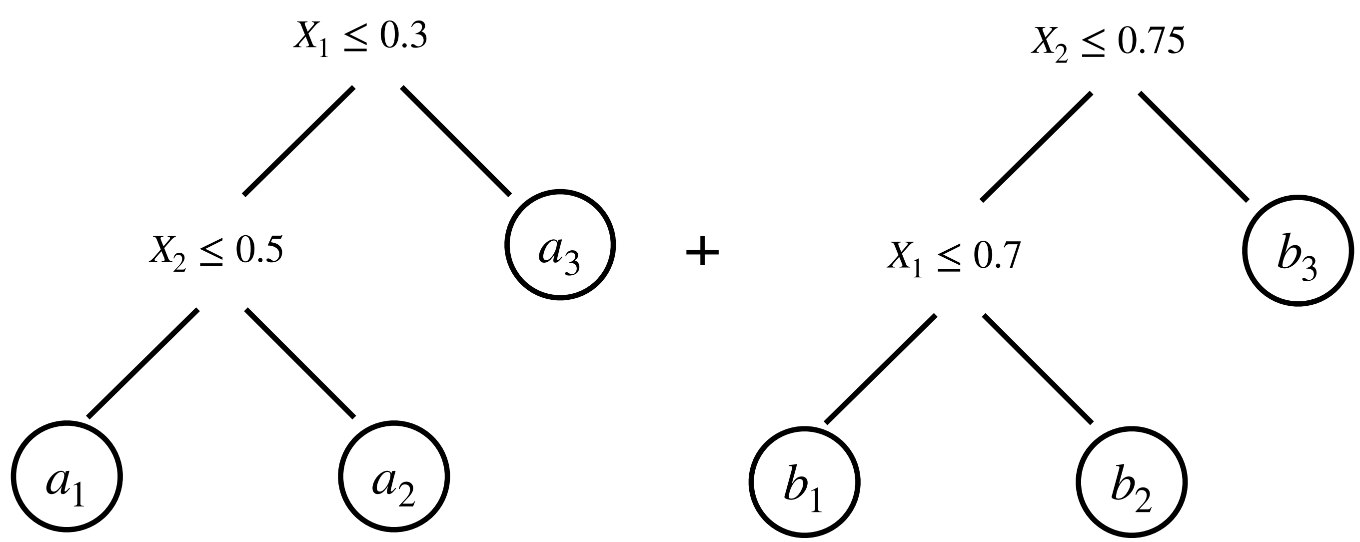
More flexibility, less overfitting!

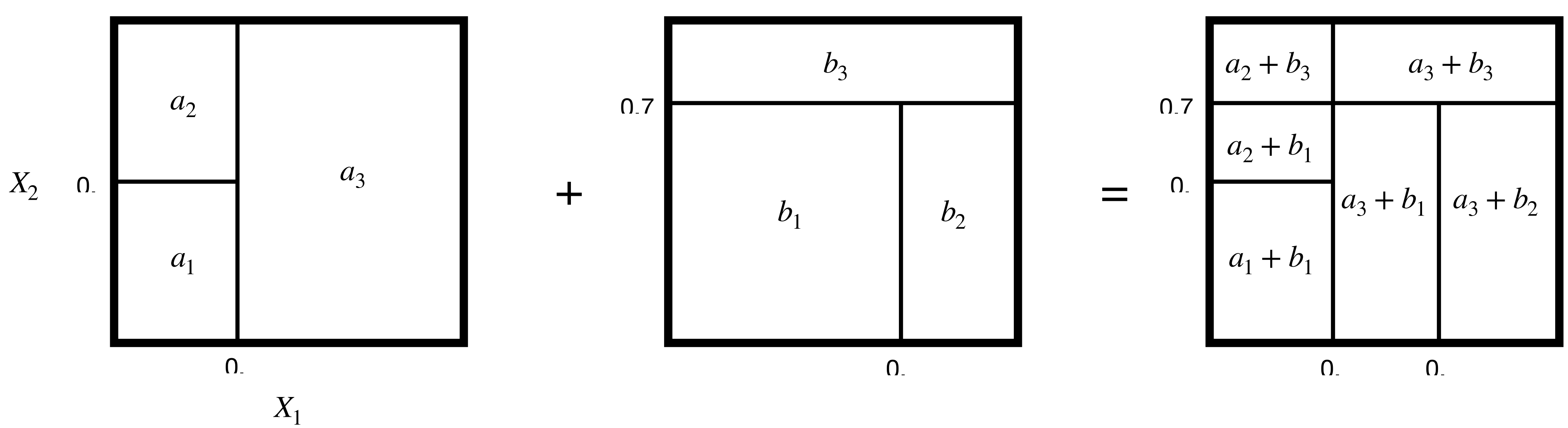


Bayesian additive regression trees

- BART isn't a model, it's a prior distribution over functions:

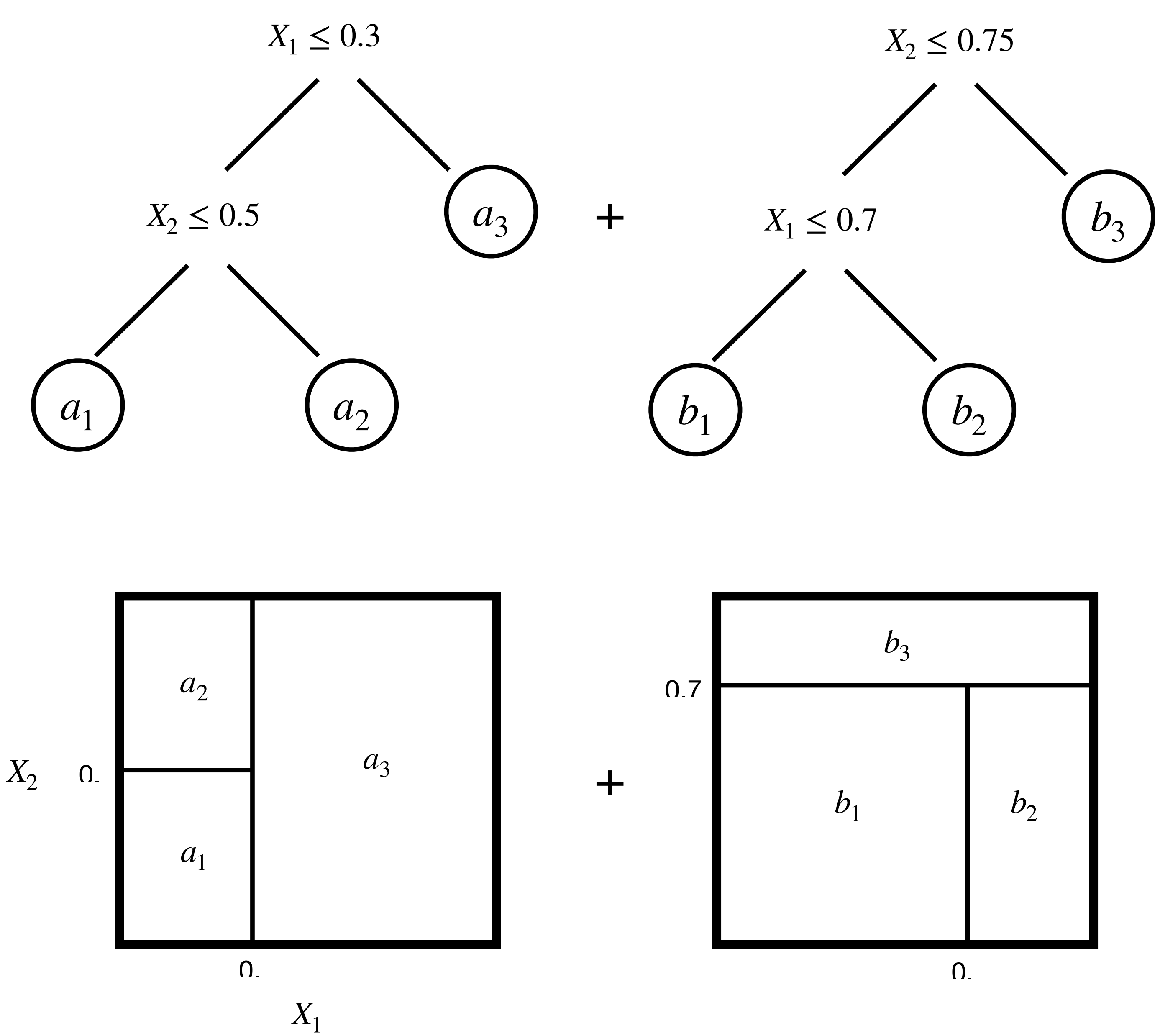
$$f(X_i) = \sum_{j=1}^{m} g(X_i, T_j, M_j)$$





The BART prior

- The prior distribution discourages (penalizes) complexity

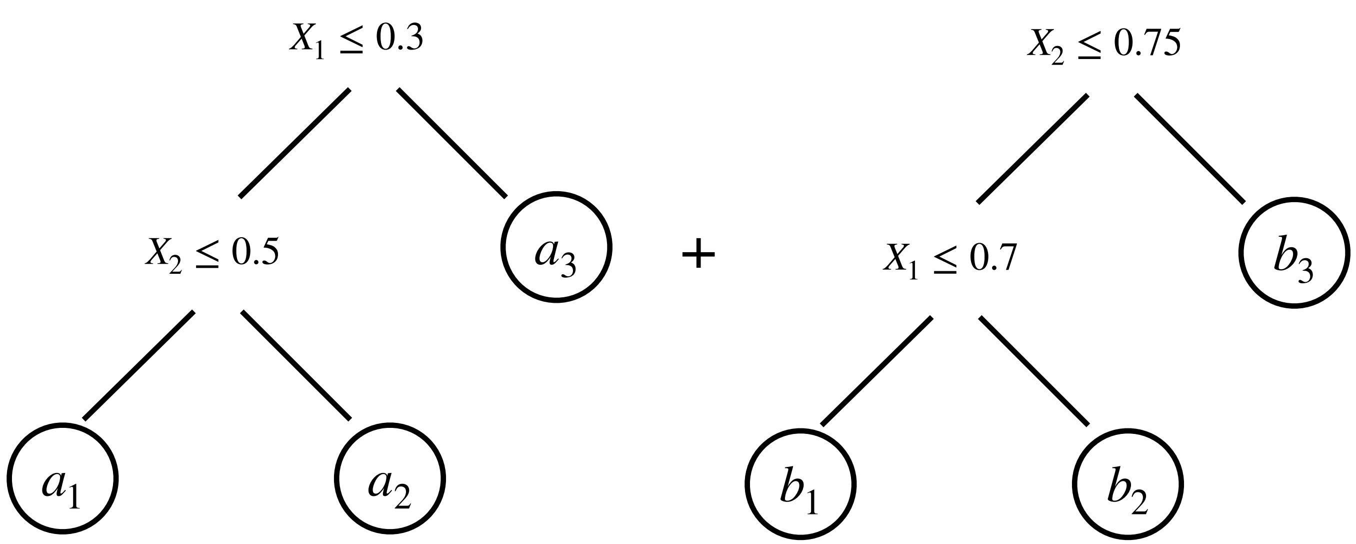


- Trees should be small (simple basis functions, few jumps)
- End-node parameters should be similar (small jumps)
- (Optionally) Splitting rules should use few variables
- The result is a relatively smooth function

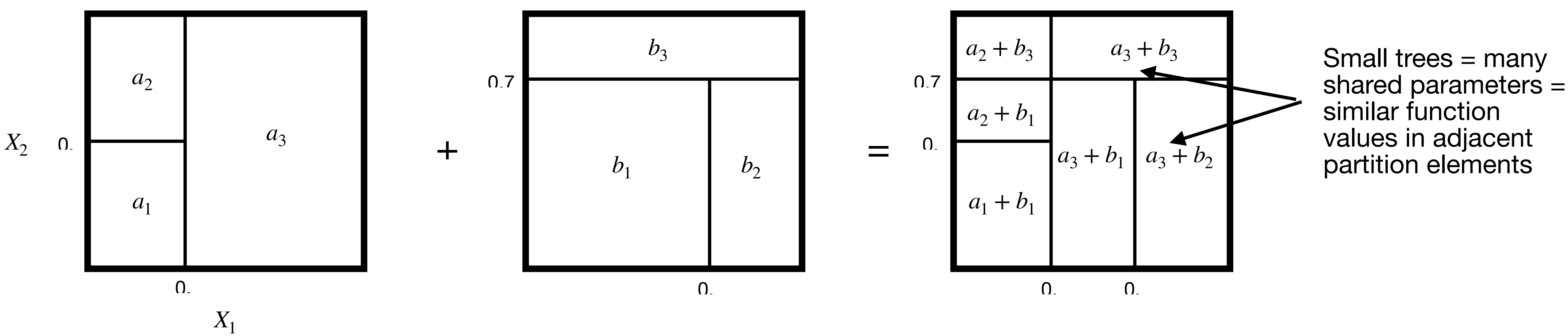
Small trees = many shared parameters = similar function values in adjacent partition elements

The BART prior

- The prior distribution discourages (penalizes) complexity



- Trees should be small (simple basis functions, few jumps)
- End-node parameters should be similar (small jumps)
- (Optionally) Splittting rules should use few variables
- The result is a relatively smooth function



BART ≈ Boosting (kind of)

Bayesian Tree Models for Causal Inference

Bayesian Tree Models for Causal Inference

- Why **Bayesian** tree models?
  - Top performers empirically
  - Robust regularization through prior distributions with sensible defaults — no “dark art” of tuning parameters
  - Inference via posterior distributions is simple and (empirically) often has good frequentist properties
  - Embedding tree methods in a *model* makes extensions straightforward — for example, to multilevel data

ACIC 2016 Data Competition

Designed to address shortcomings of existing comparisons:

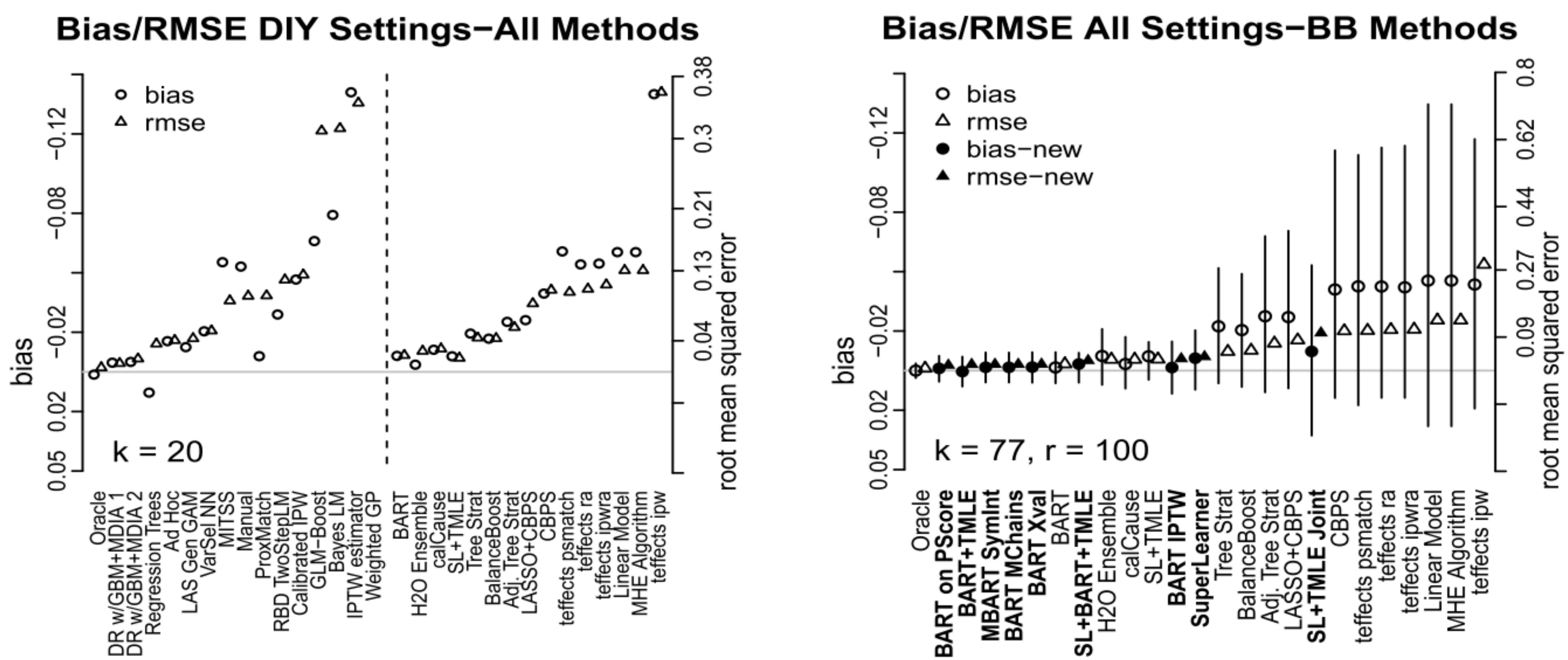
1. Few methods compared and unfair comparisons.
2. Testing grounds not calibrated to "real life."
3. File drawer effect(s)

"Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition", Dorie et al. (2019)

Two tracks: Automated vs. “Do-It-Yourself”

30 different competing methods

77 different “true worlds” where data were simulated under different settings of nonlinearity, % treated, overlap, alignment, treatment heterogeneity, treatment magnitude — 100 datasets per world = 7,700 total datasets



“Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition”, Dorie et al. (2019)

“Only a subset of the methods output individual level treatment effect estimates… Of these, the BART methods and calCause performed noticeably better than the other options; however, since the two other competitors were quite simple and relied on linear models this was not particularly surprising”

Table 4: Abbreviated ACIC 2016 contest results. Coverage and average interval length are reported for nominal 95% uncertainty intervals. Bias and |Bias| are average bias and average absolute bias, respectively, over the. PEHE is the average precision in estimating heterogeneous treatment effects (the average root mean squared error of CATE estimates for each unit in a dataset) (Hill, 2011).

	Coverage	Int. Len.	Bias	(SD)	|Bias|	(SD)	PEHE	(SD)
BCF	0.82	0.026	-0.0009	(0.01)	0.008	0.010	0.33	0.18
ps-BART	0.88	0.038	-0.0011	(0.01)	0.010	0.011	0.34	0.16
BART	0.81	0.040	-0.0016	(0.02)	0.012	0.013	0.36	0.19
Causal RF	0.58	0.055	-0.0155	(0.04)	0.029	0.027	0.45	0.21

Table 5: Tests and estimates for differences between BCF and other methods in the ACIC 2016 competition. The p-values are from bootstrapp permutation tests with 100,000 replicates.

	Diff Bias	p	Diff |Bias|	p	Diff PEHE	p
ps-BART	-0.00020	0.146	0.0011	$< 1e^{-4}$	0.010	$< 1e^{-4}$
BART	-0.00070	$< 1e^{-4}$	0.0031	$< 1e^{-4}$	0.037	$< 1e^{-4}$
Causal RF	-0.01453	$< 1e^{-4}$	0.0204	$< 1e^{-4}$	0.125	$< 1e^{-4}$

Hahn, P. Richard, Jared S. Murray, and Carlos M. Carvalho. "Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion)." Bayesian Analysis 15.3 (2020): 965-1056.

ACIC 2017 Competition

- Richard Hahn, Vincent Dorie, and myself
- Considered non-additive or heterosketastic errors, extra replicated datasets, **more realistic effect sizes/heterogeneity** and **explicit models of confounding**
- Substantively similar results — Bayesian tree models work well, most methods aren't built for heterogeneity, semiparametric adjustments are useful (sometimes).

Hahn, P. Richard, Vincent Dorie, and Jared S. Murray. "Atlantic Causal Inference Conference (ACIC) Data Analysis Challenge 2017." *arXiv preprint arXiv:1905.09515* (2019).

Bayesian Causal Forests in Education

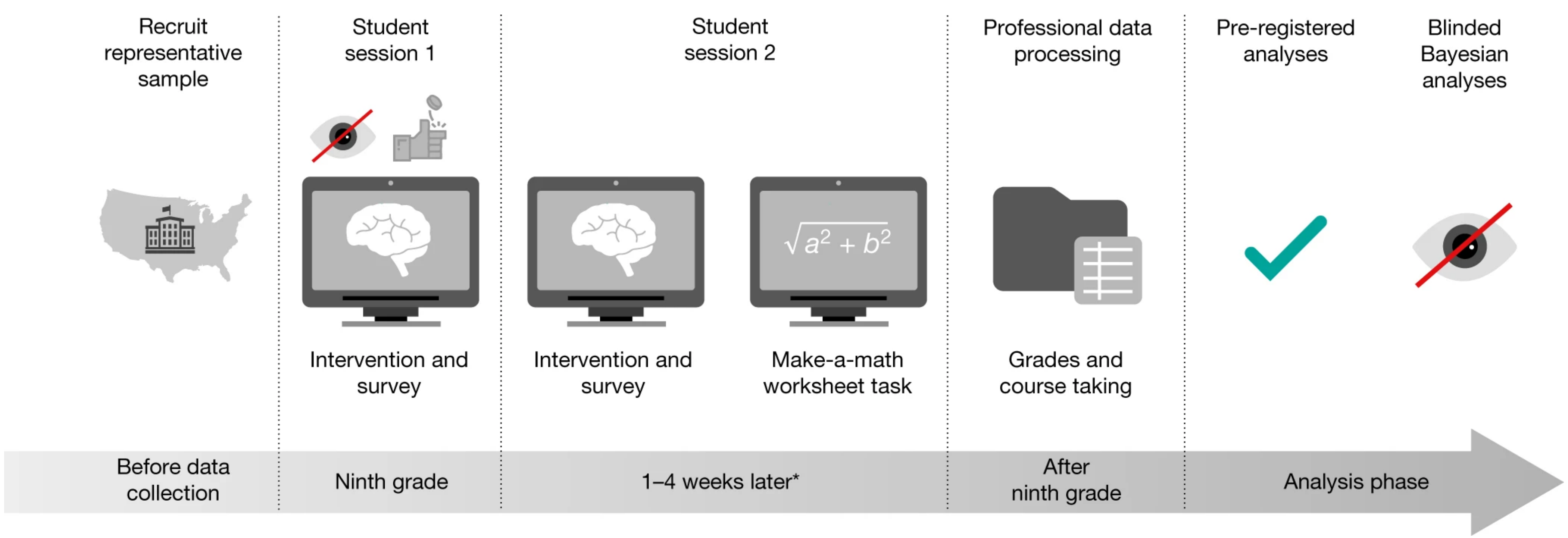
Article | Open access | Published: 07 August 2019

A national experiment reveals where a growth mindset improves achievement

David S. Yeager, Paul Hanselman, Gregory M. Walton, Jared S. Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Chris S. Hulleman, Cintia P. Hinojosa, David Paunesku, Carissa Romero, Kate Flint, Alice Roberts, Jill Trott, Ronaldo Iachan, Jenny Buontempo, Sophia Man Yang, Carlos M. Carvalho, P. Richard Hahn, Maithreyi Gopalan, Pratik Mhatre, Ronald Ferguson, Angela L. Duckworth & Carol S. Dweck

Nature 573, 364–369 (2019) | Cite this article

BCF analysis



A simple multilevel linear model

Group-specific intercepts/
fixed/random effects

Group-specific “unexplained”
heterogeneity

$$y_{ij} = \alpha_j + \mathbf{x}'_{ij}\beta + \left[\mathbf{w}'_{ij}\tau + \phi_j\right] z_{ij} + \epsilon_{ij}$$

Controls at the individual
and/or group level

Moderators at the individual
and/or group level

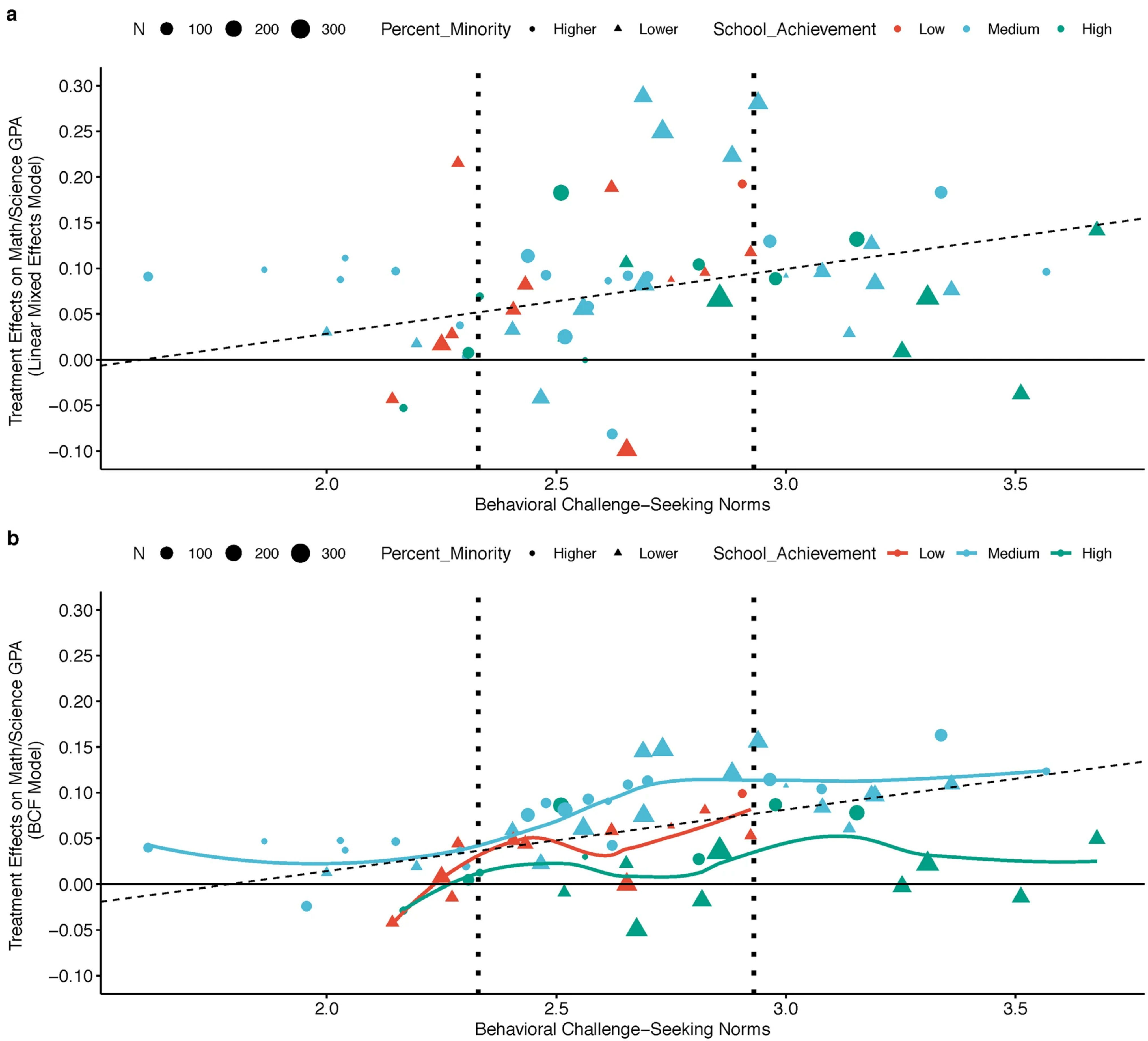
Multilevel Bayesian Causal Forests

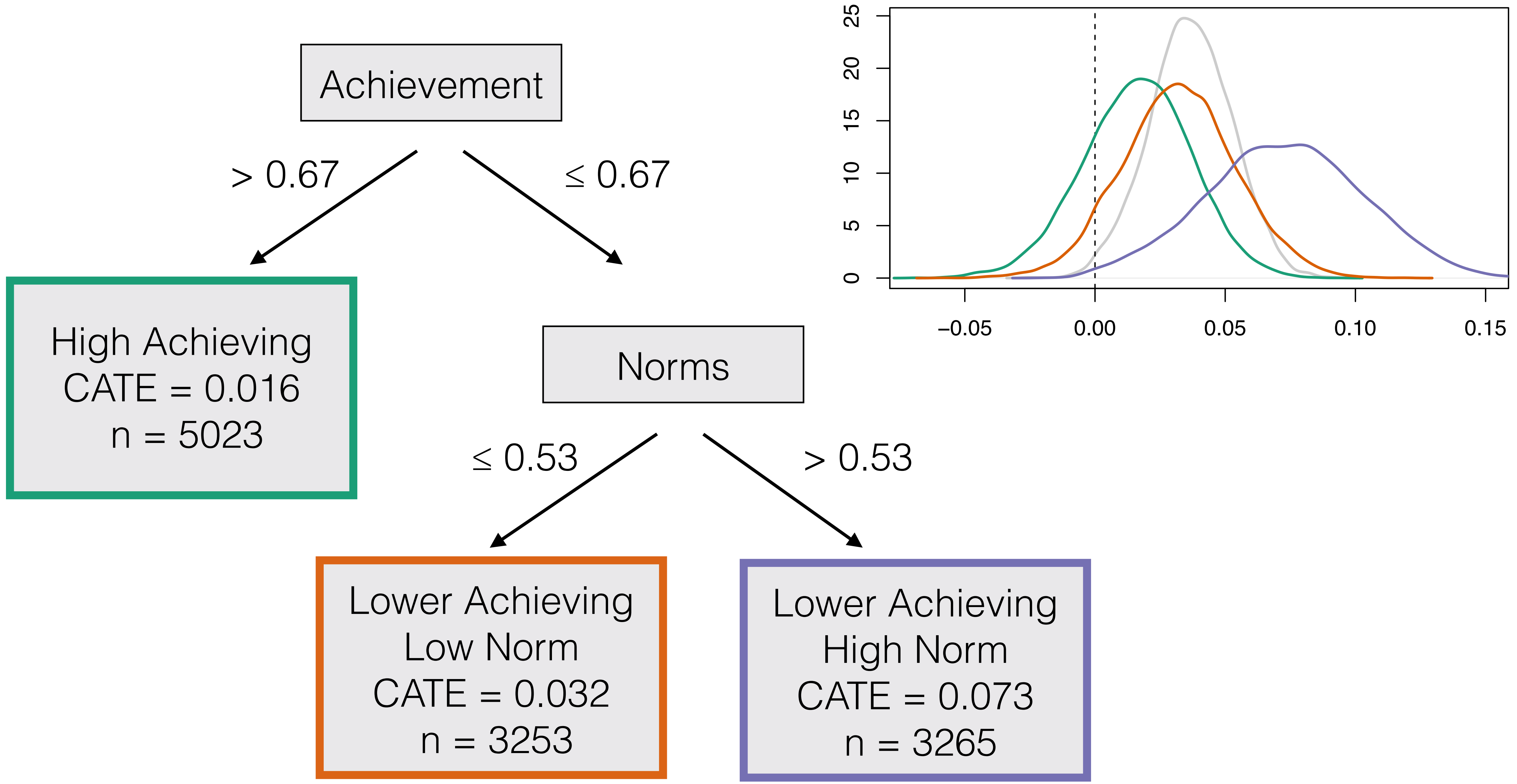
$$y_{ij} = \alpha_j + \beta(\mathbf{x}_{ij}) + \left[\tau(\mathbf{w}_{ij}) + \phi_j\right] z_{ij} + \epsilon_{ij}$$

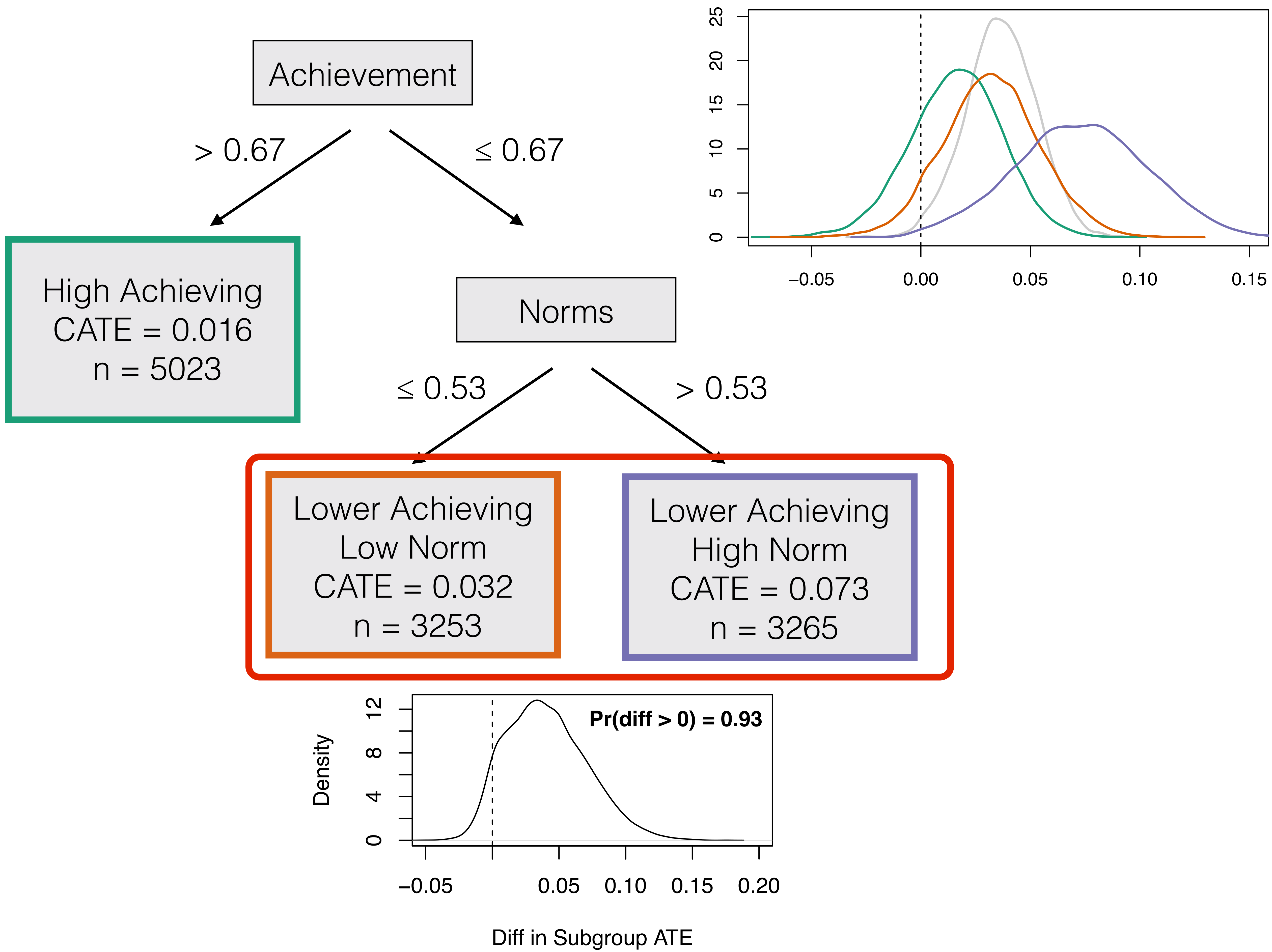
Replace linear terms
with BART

BART in causal inferece: Hill (2011), Green & Kern (2012), …

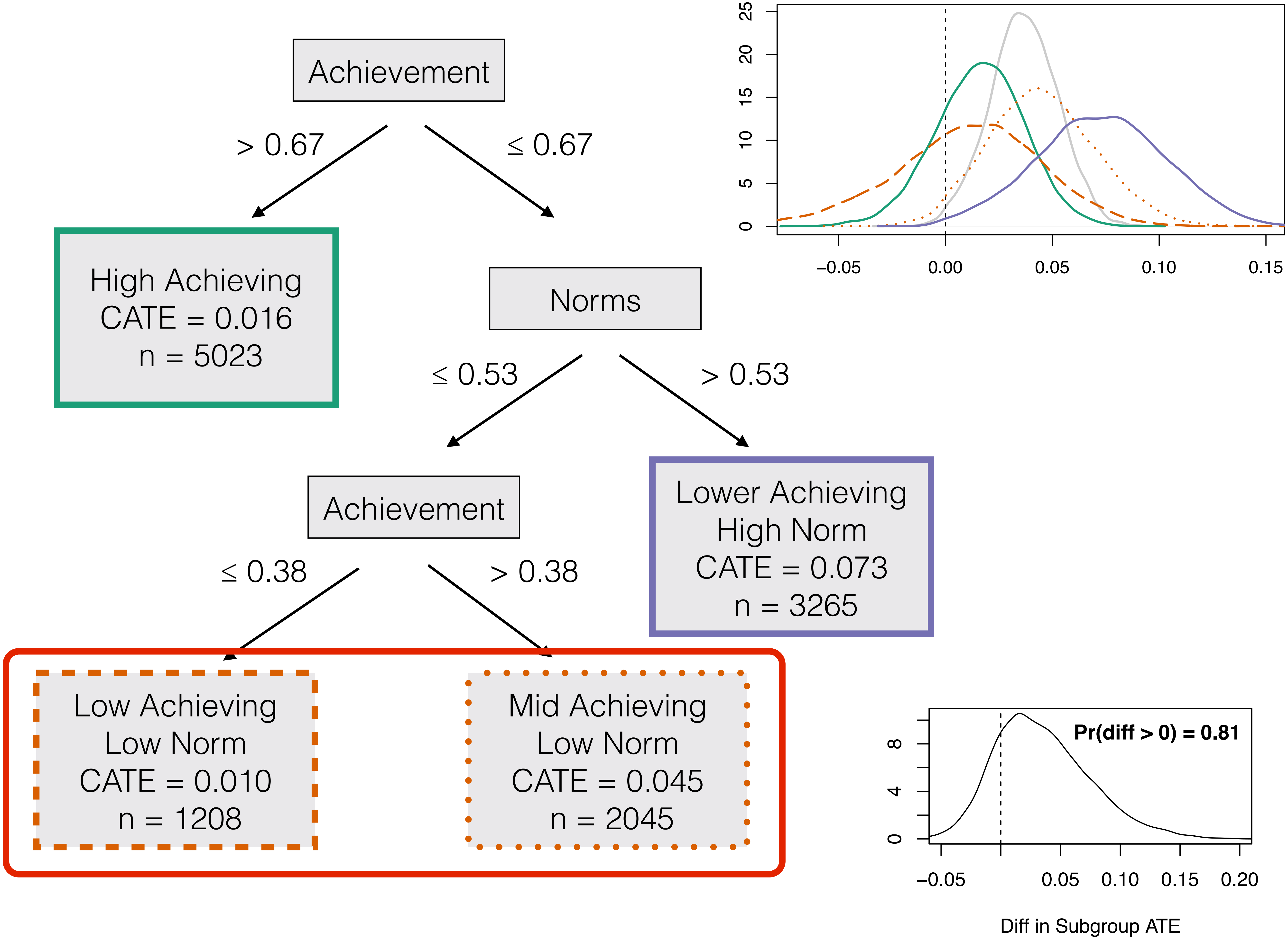
Bayesian Causal Forests: Parameterizing treatment effect heterogeneity with BART from Hahn, Murray and Carvalho (2020)


a
N 100 200 300
Percent_Minority Higher Lower
School_Achievement Low Medium High
Treatment Effects on Math/Science GPA (Linear Mixed Effects Model)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
2.0
2.5
3.0
3.5
Behavioral Challenge−Seeking Norms
b
N 100 200 300
Percent_Minority Higher Lower
School_Achievement Low Medium High
Treatment Effects on Math/Science GPA (BCF Model)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
2.0
2.5
3.0
3.5
Behavioral Challenge−Seeking Norms


Achievement
> 0.67
≤ 0.67
High Achieving
CATE = 0.016
n = 5023
Norms
≤ 0.53
> 0.53
Lower Achieving
Low Norm
CATE = 0.032
n = 3253
Lower Achieving
High Norm
CATE = 0.073
n = 3265
0
5
10
15
20
25
−0.05
0.00
0.05
0.10
0.15



Achievement
> 0.67
≤ 0.67
High Achieving
CATE = 0.016
n = 5023
Norms
≤ 0.53
> 0.53
Lower Achieving
Low Norm
CATE = 0.032
n = 3253
Lower Achieving
High Norm
CATE = 0.073
n = 3265
0
5
10
15
20
25
−0.05
0.00
0.05
0.10
0.15
Pr(diff > 0) = 0.93
Density
0
4
8
12
−0.05
0.05
0.10
0.15
0.20
Diff in Subgroup ATE

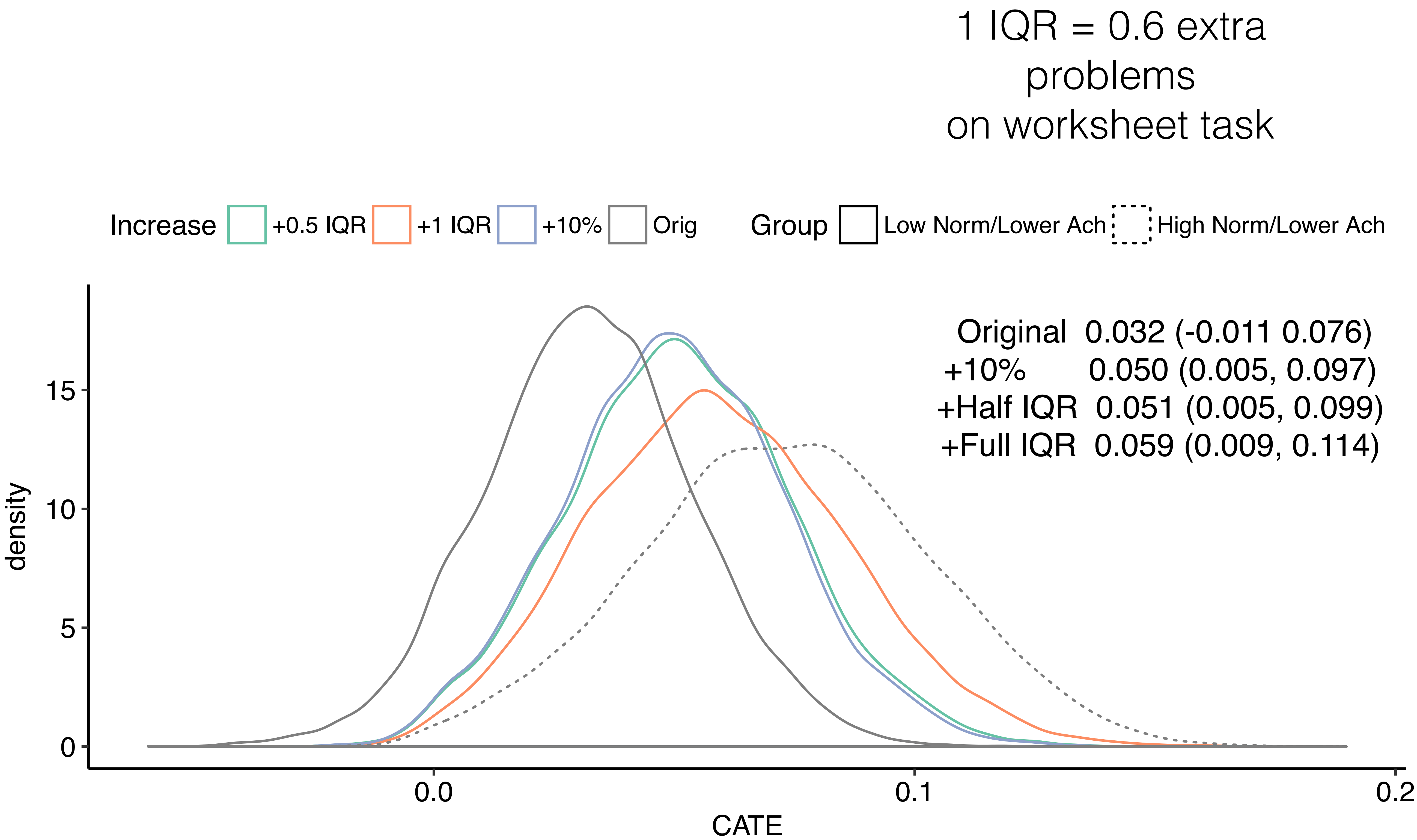


Achievement
> 0.67
≤ 0.67
High Achieving
CATE = 0.016
n = 5023
Norms
≤ 0.53
> 0.53
Achievement
≤ 0.38
> 0.38
Lower Achieving
High Norm
CATE = 0.073
n = 3265
Low Achieving
Low Norm
CATE = 0.010
n = 1208
Mid Achieving
Low Norm
CATE = 0.045
n = 2045
Pr(diff > 0) = 0.81
Diff in Subgroup ATE

Counterfactual treatment effect predictions

- How do estimated treatment effects change in lower achieving/ low norm schools if norms increase, holding constant school minority comp & achievement?

- Not a formal causal mediation/moderation analysis (roughly, we would need “no unmeasured moderators correlated with norms”)



1 IQR = 0.6 extra
problems
on worksheet task
Increase
+0.5 IQR
+1 IQR
+10%
Orig
Group
Low Norm/Lower Ach
High Norm/Lower Ach
Original 0.032 (-0.011 0.076)
+10% 0.050 (0.005, 0.097)
+Half IQR 0.051 (0.005, 0.099)
+Full IQR 0.059 (0.009, 0.114)
density
15
10
5
0
0.0
0.1
0.2
CATE

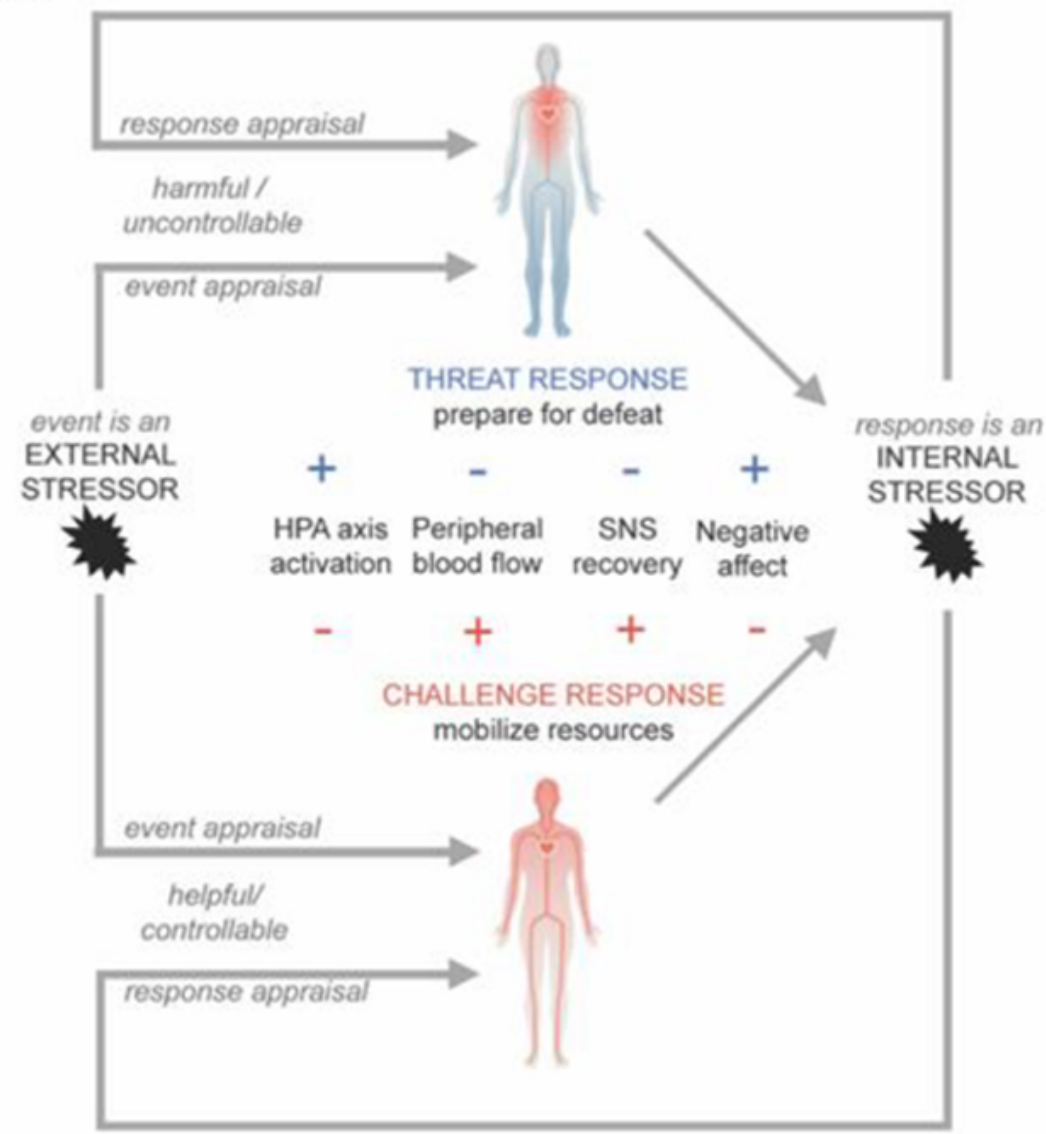
Let's see a demo!

Article

A Synergistic Mindsets Intervention Protects Adolescents from Social-Evaluative Stress

David S. Yeager[1*], Christopher J. Bryan[1*], James J. Gross[2], Jared Murray[1], Danielle Krettek[3], Pedro Santos[1], Hannah Graveling[4], Meghann Johnson[1], Jeremy P. Jamieson[4*]

Studies (Sample Size)	Population	Stressor	Measures of Threat-Type Stress Response
1 (N = 2,717)	13-18 y/o U.S. public school students during the COVID-19 pandemic.	Anticipated timed assignment	Event- and response-focused appraisals
2 (N = 755)	Diverse undergraduate students attending a public university	Experienced timed assignment	Cognitive appraisals at 1-3 days and 3 weeks post-test
3 and 4 (3: N = 160; 4: N = 200)	Undergraduate students at a private university	Trier Social Stress Test	Peripheral blood flow
5 (N = 118, n=1213 observations)	14-16 y/o adolescents from racial/ethnic minority groups, facing economic disadvantages	Daily stressors in high school	Daily negative self-regard and HPA-axis activation
6 (N = 341)	Same as study 2 but during the onset of the COVID-19 pandemic in Spring 2020	Ongoing academic demands during COVID-19 quarantines	Generalized internalizing symptoms



The Synergistic Mindsets Intervention

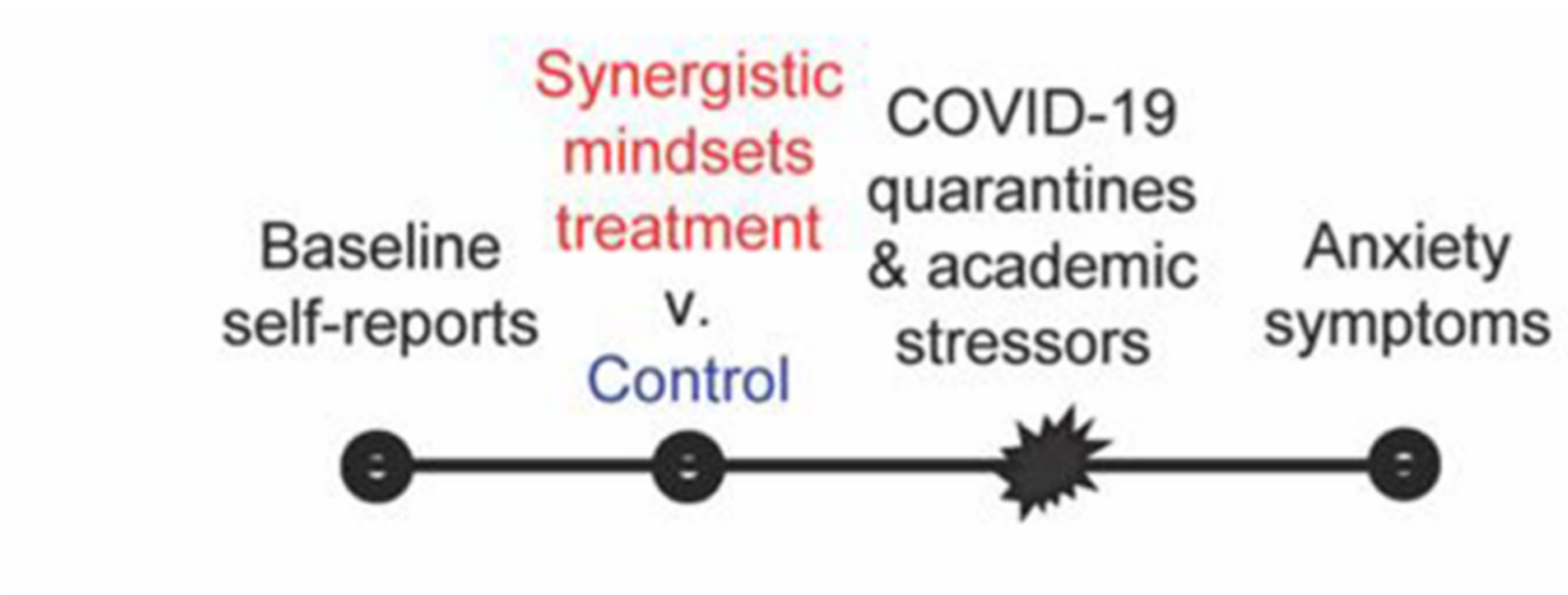
Growth mindset	+	**Stress-can-be-enhancing mindset**
e.g. Yeager & Dweck (2012)		e.g. Crum et al. (2013)

- *Your abilities can be developed;*
- Understanding that helps you see struggles as a path to improvement (not a sign of inability).

- *Your stress responses can be useful for enhancing performance;*
- Understanding that allows you to harness the stress responses that inevitably follow from challenging goal-pursuit.

Expected intervention effects would be moderated by pre-intervention mindsets (largest effects in poor-mindset students) and the strength of the stressor

Study 6: Anxiety symptoms post-COVID-19



End of
Jan
2020

Mid
Apr
2020

Thank you!