

# The Holy Trinity: Blending Statistics, Machine Learning and Discrete Choice with Applications to Strategic Bicycle Planning

Timothy Brathwaite

UC Berkeley

Civil and Environmental Engineering

# More Cars, More Problems

VIDEO: 2 deadly traffic accidents on Highway 880, 1 in San Francisco Monday morning

By Julianne Herrera, KRON

Published: May 1, 2017, 6:15 am | Updated: May 1, 2017, 6:42 am



## WEATHER

BAY AREA AMONG MOST POLLUTED REGIONS IN THE COUNTRY

[Share](#) [G+1](#) [Tweet](#) [Email](#)



The San Francisco Bay Area continues to struggle with air quality. (KGO-TV)

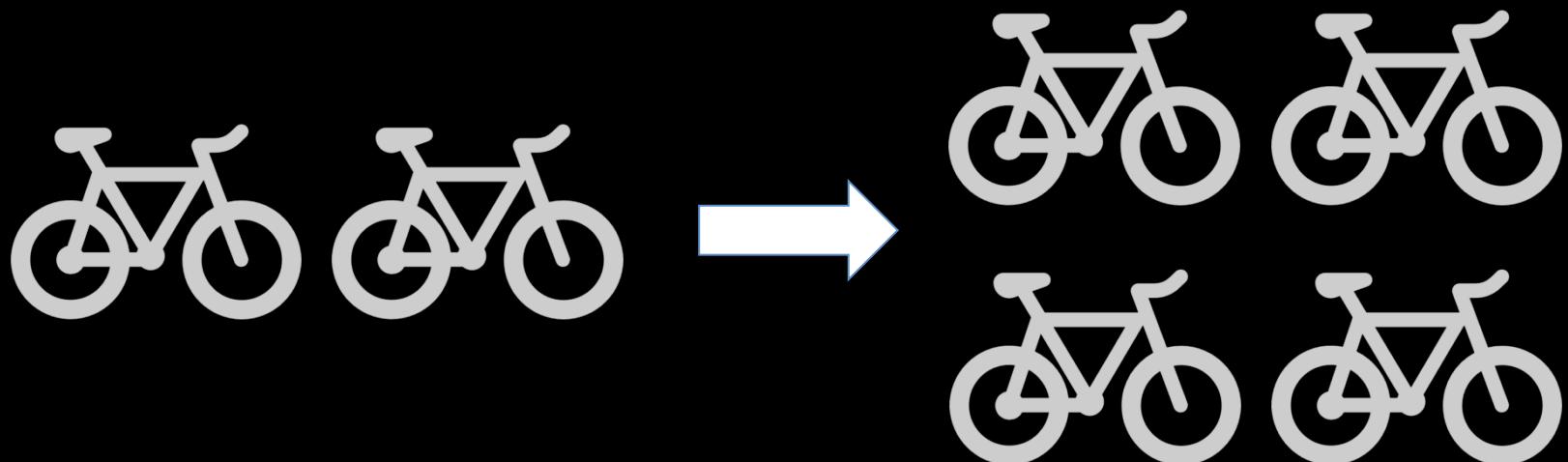


Wednesday, April 19, 2017



# One Governmental Response:

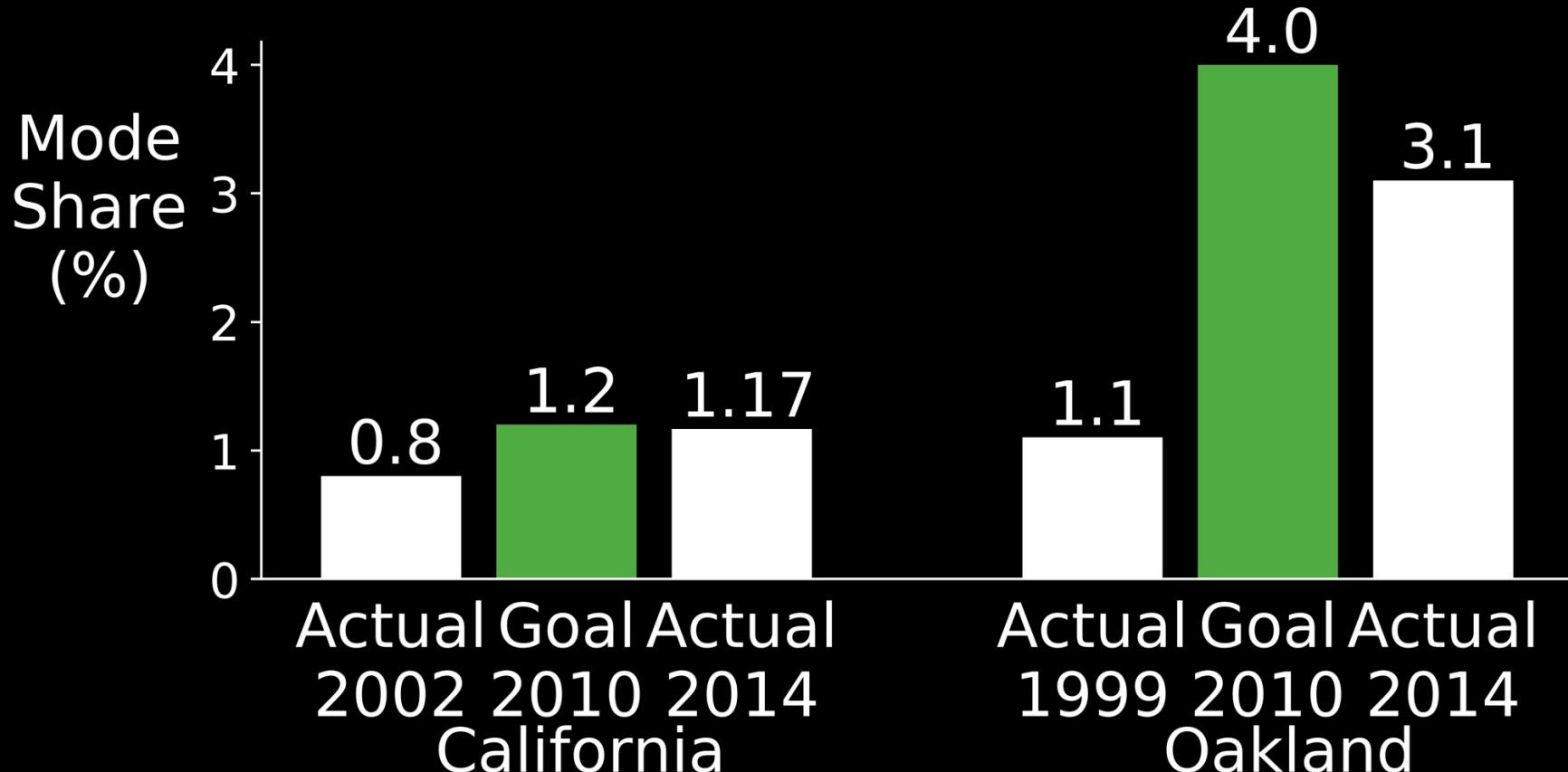
Promote increased bicycle usage.



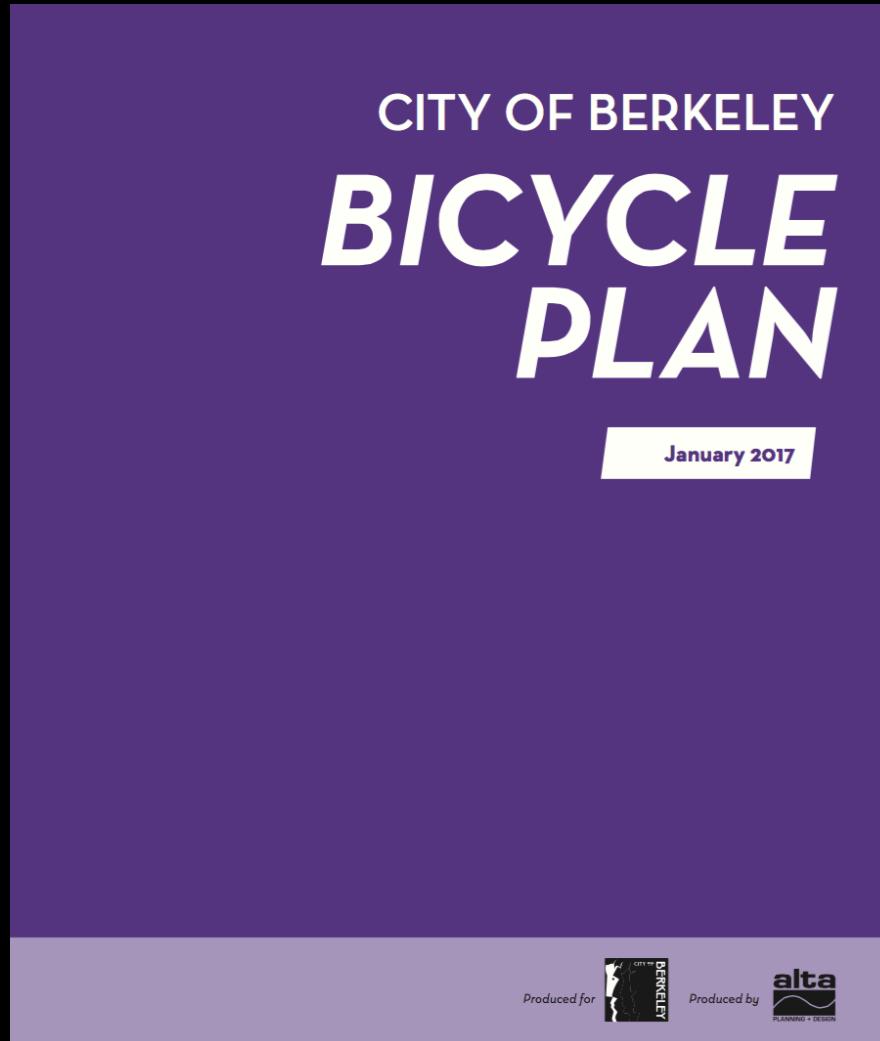
# Outcomes have been poor:

Bicycle mode share goals are routinely missed.

Actual and Goal California Commute Mode Shares



# Limited funding necessitates strategic investments



# Limited funding necessitates strategic investments

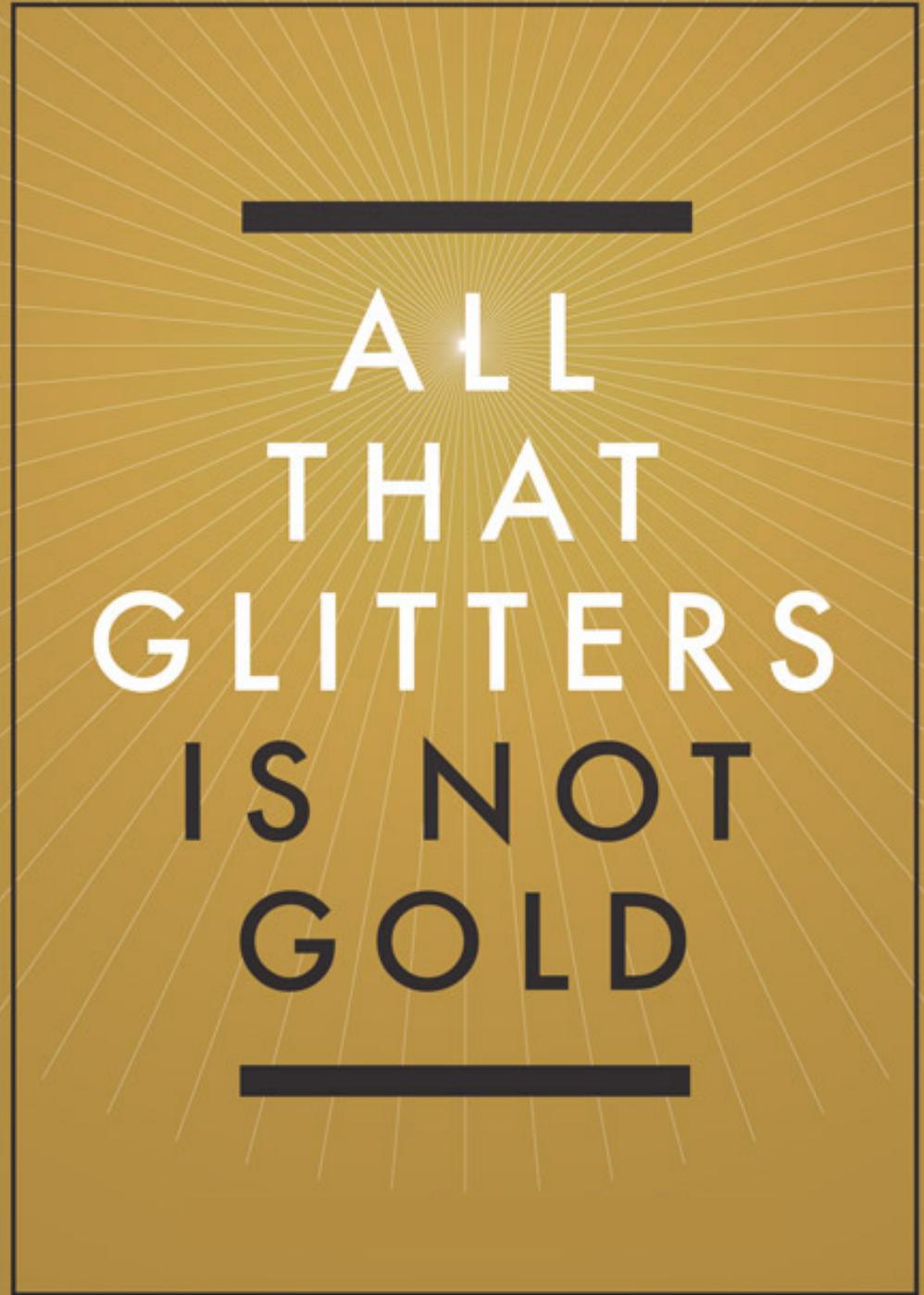
**Table 6-1: Evaluation Criteria**

CRITERIA	DESCRIPTION	MAX SCORE
Safety	Combination of safety, LTS, and demand analysis	50
Community Support	Projects are scored based on whether the project or area was identified for improvement during the initial community input phase	30
Equity	Projects are scored based on whether they are located within a MTC designated Community of Concern.	20
	Total Possible Score	100

How best to evaluate effects on  
bicycle demand?

**Gold Standard: Travel Demand Model**





—  
ALL  
THAT  
GLITTERS  
IS NOT  
GOLD  
—

**WHO WILL TELL  
THE EMPEROR HE  
HAS NO CLOTHES?**

# What are (some of) the problems?

- Omission of relevant variables
- Use of questionable behavioral assumptions
- Ignoring statistical implications of the small number of cyclists.

# My Contributions

- “Zone of Likely Travel” + Decision Tree
  - Omission of relevant variables
- Tree-Based Gaussian Process Choice Model
  - Questionable behavioral assumptions
- Asymmetric Choice Models
  - Statistical implications of the small # of cyclists.

# Sí, se puede

Increasing the relevance of travel demand models for bicycle planning simultaneously increases their accuracy.



# Outline

1. What's a model?
2. How to incorporate bike relevant variables?
3. How to model “irrational” decision making?
4. How to model the “minority” mode?
5. Recap: So what?
6. What next?

# Anatomy of a Model

## Inputs

- Socio-demographics
- Level of Service  
(e.g. Travel Time / Cost)

## Outputs

- Probability of traveling by each mode



# What's missing?

## Inputs

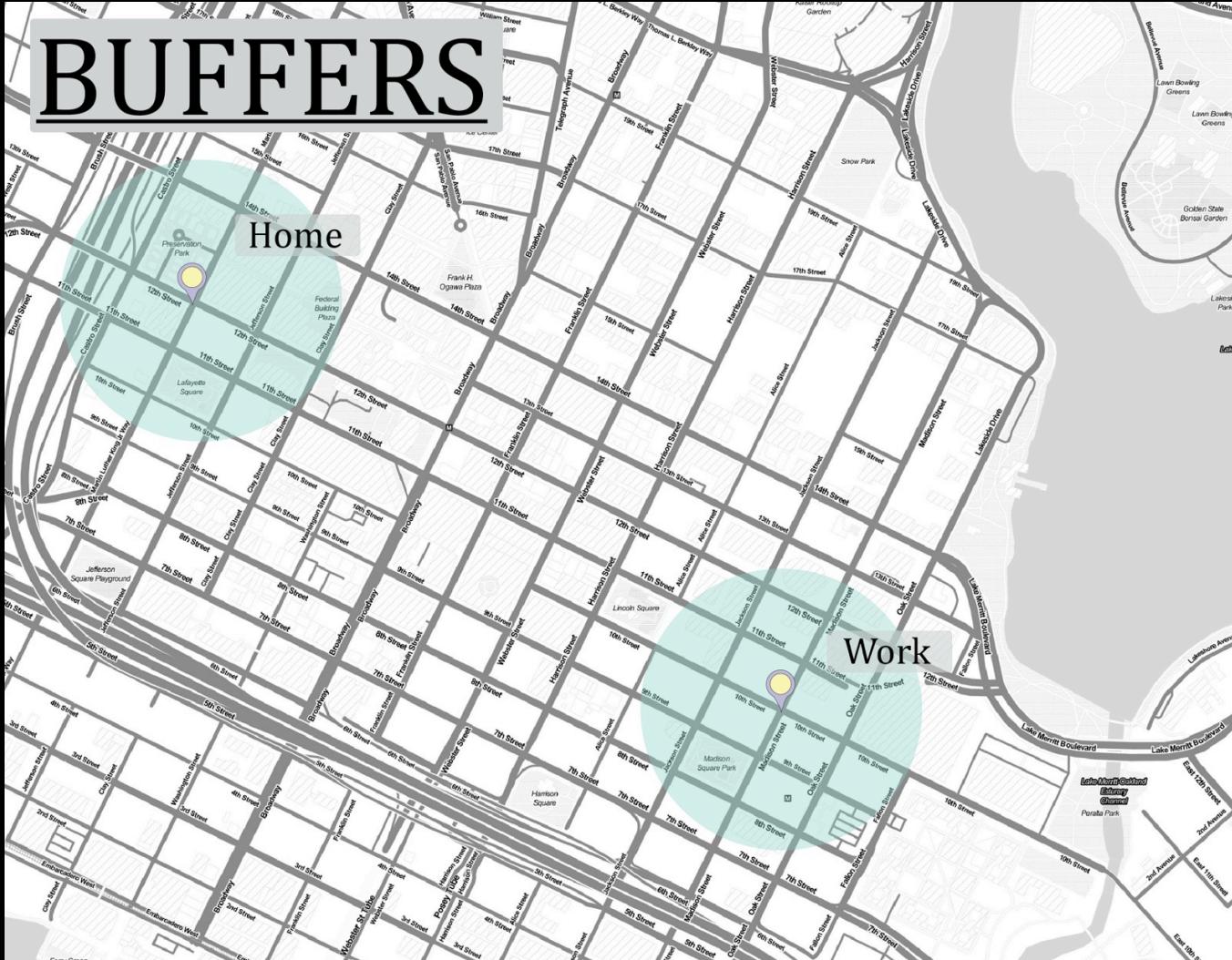
- Socio-demographics
- Level of Service  
(e.g. Travel Time / Cost)
- Bike Infrastructure?  
(e.g. lanes, paths, signs)
- Traffic volumes?
- Traffic speeds?
- Topography?

## Outputs

- Probability of traveling by each mode



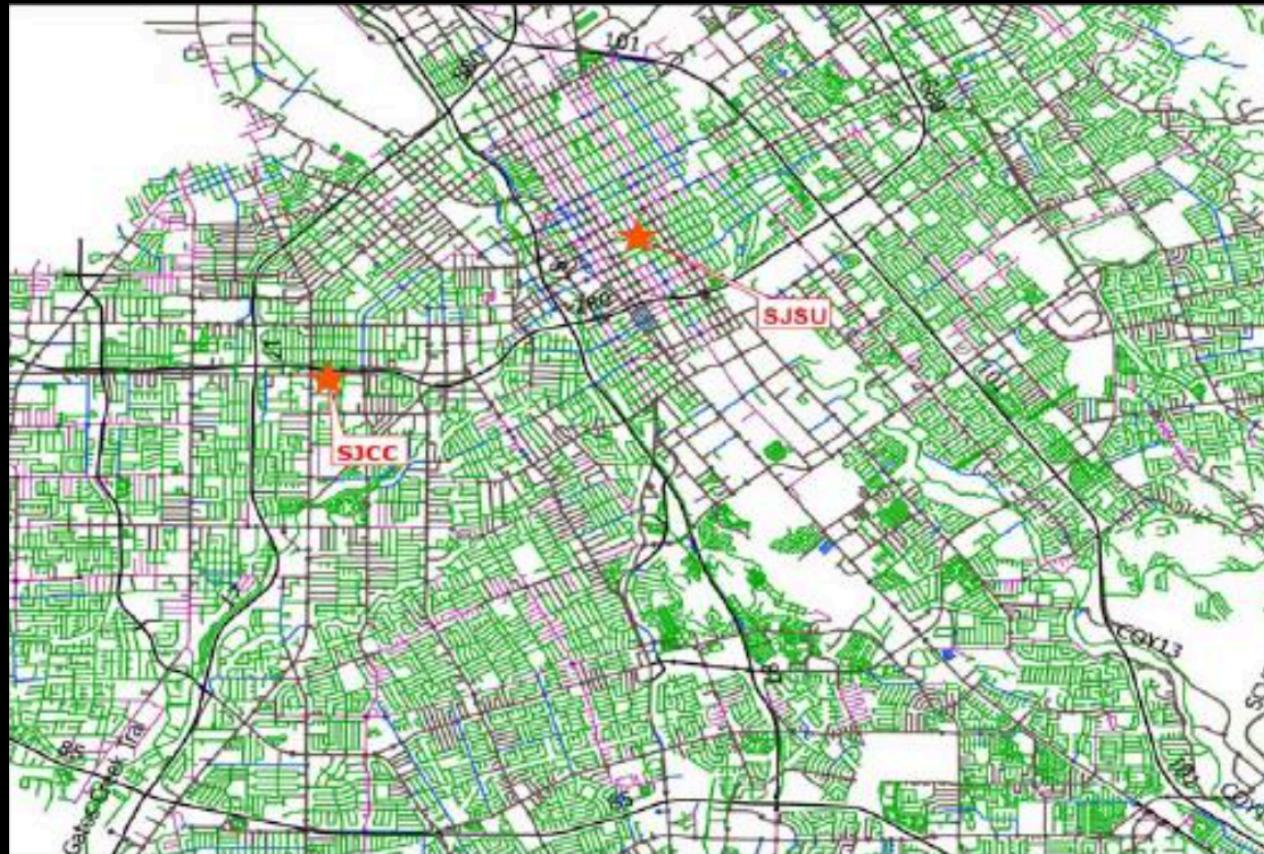
# How have roadway level variables been previously incorporated?



# How have roadway level variables been previously incorporated?

Bicycle  
Environment  
Factors:

1. Score roadways
2. Aggregate scores

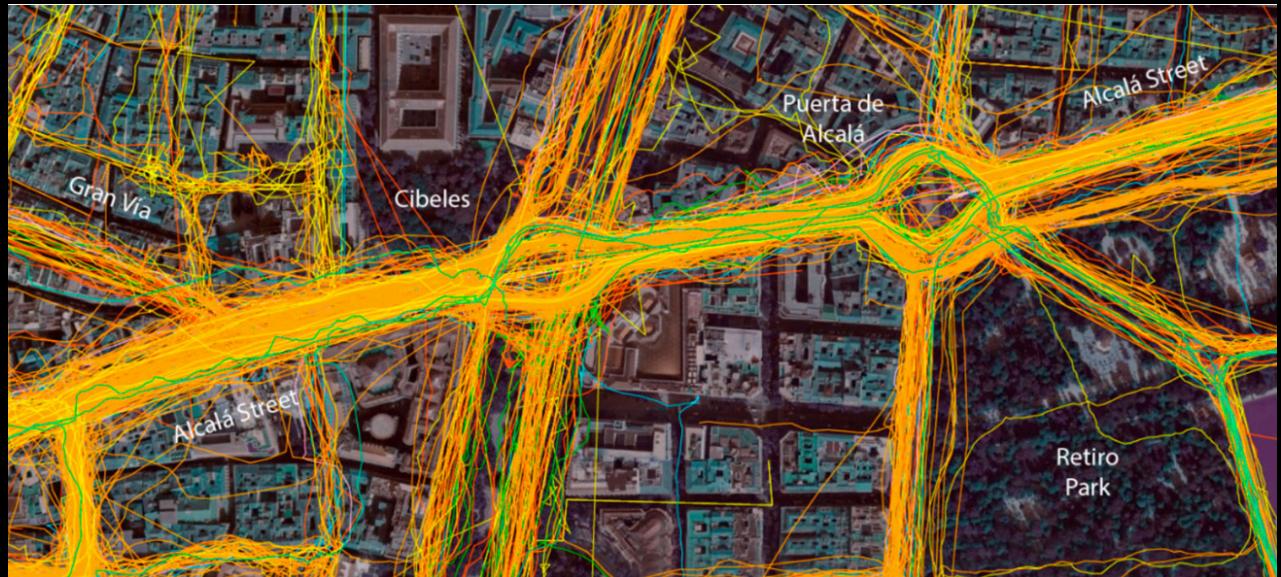


# How have roadway level variables been previously incorporated?

## Route Choice

### Models:

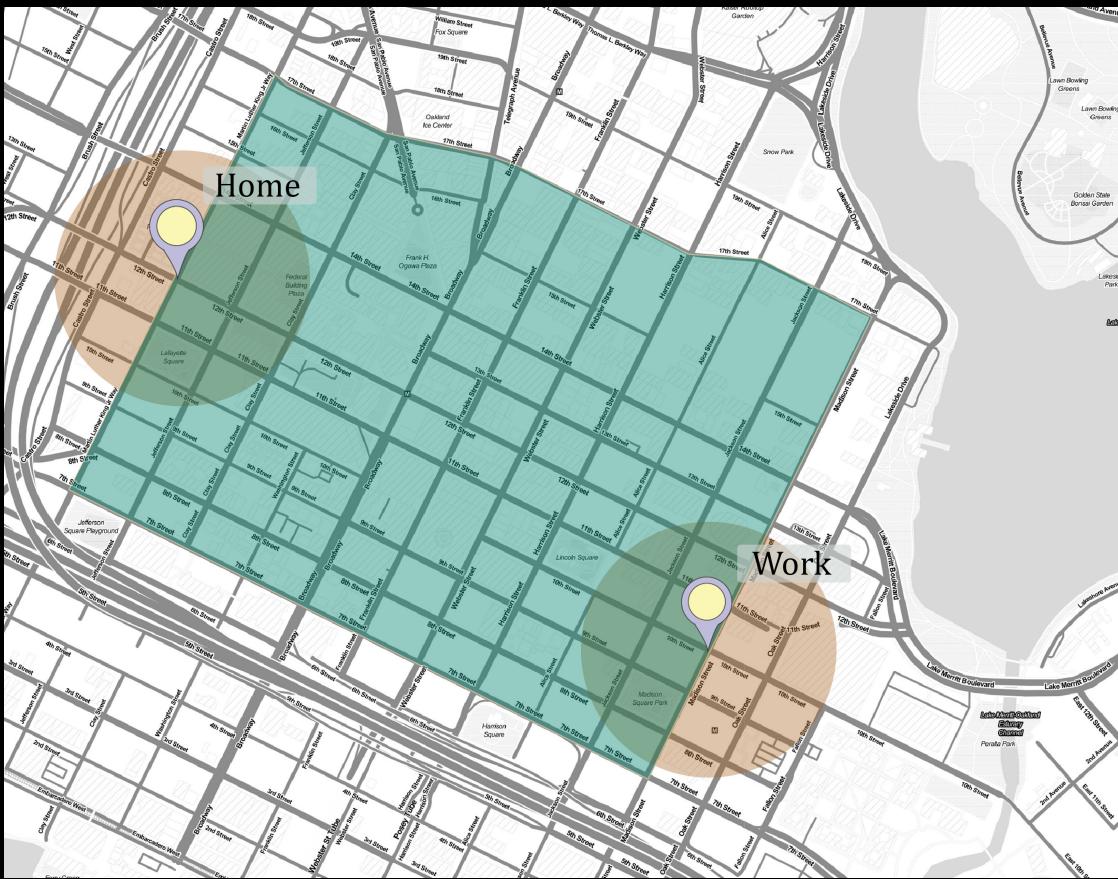
1. Score routes
2. Aggregate scores



Romanillos and Austwick (2016), "Madrid Cycle Track: Visualizing the Cyclable City"  
Hood et al. (2011), "A GPS-based bicycle route choice model for San Francisco"

# Problems with past approaches?

Buffers include too many irrelevant streets and exclude too many relevant streets.



# Problems with past approaches?

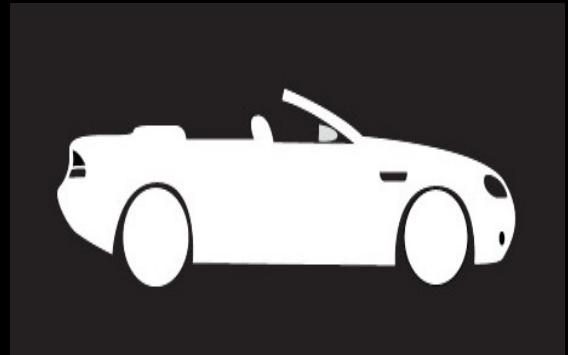
Bicycle environment factors are either completely subjective or “un-interpretable.”



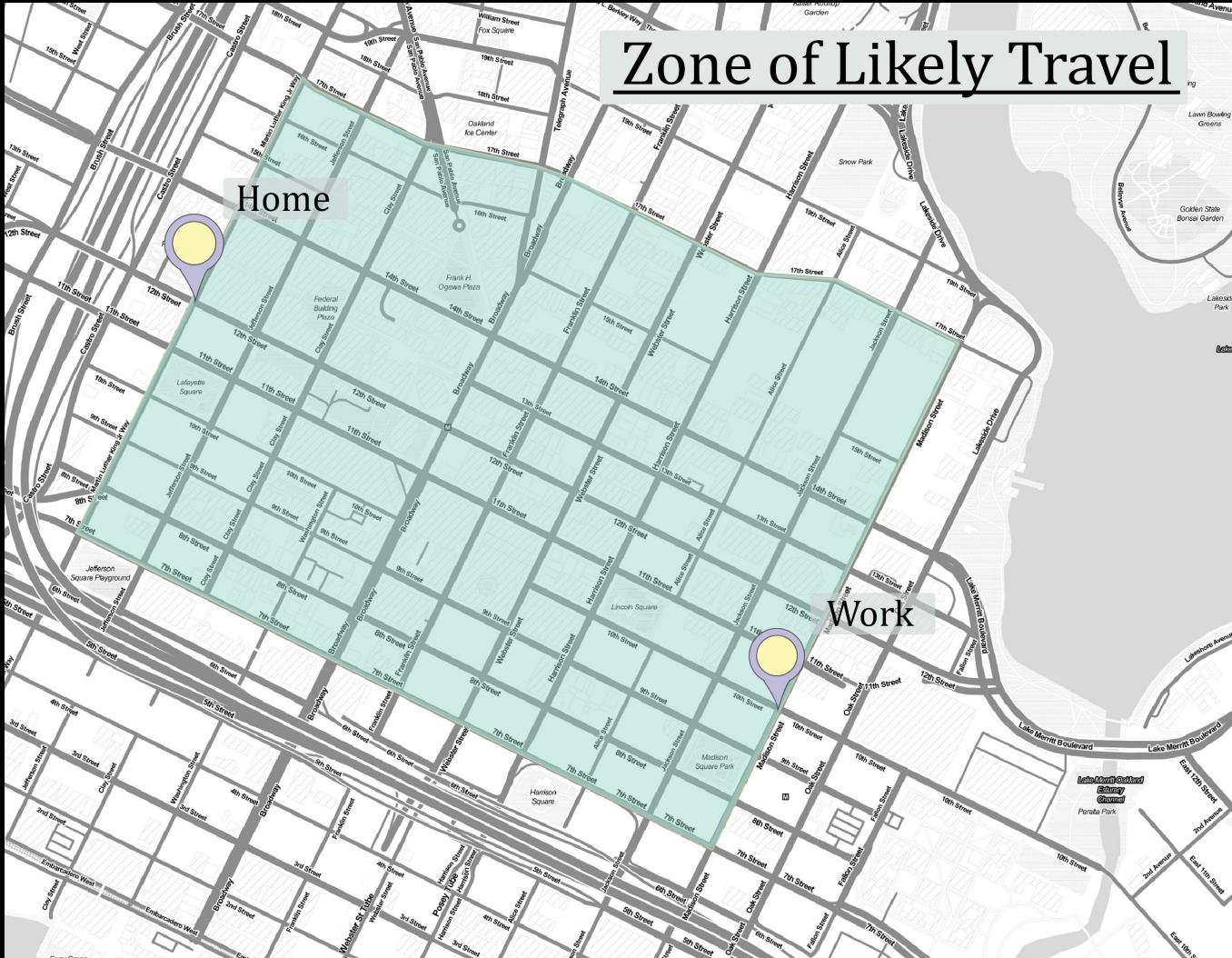
“Factor Analysis”

# Problems with past approaches?

Route Choice Models implicitly assume that current cyclists and non-cyclists have equal sensitivities.

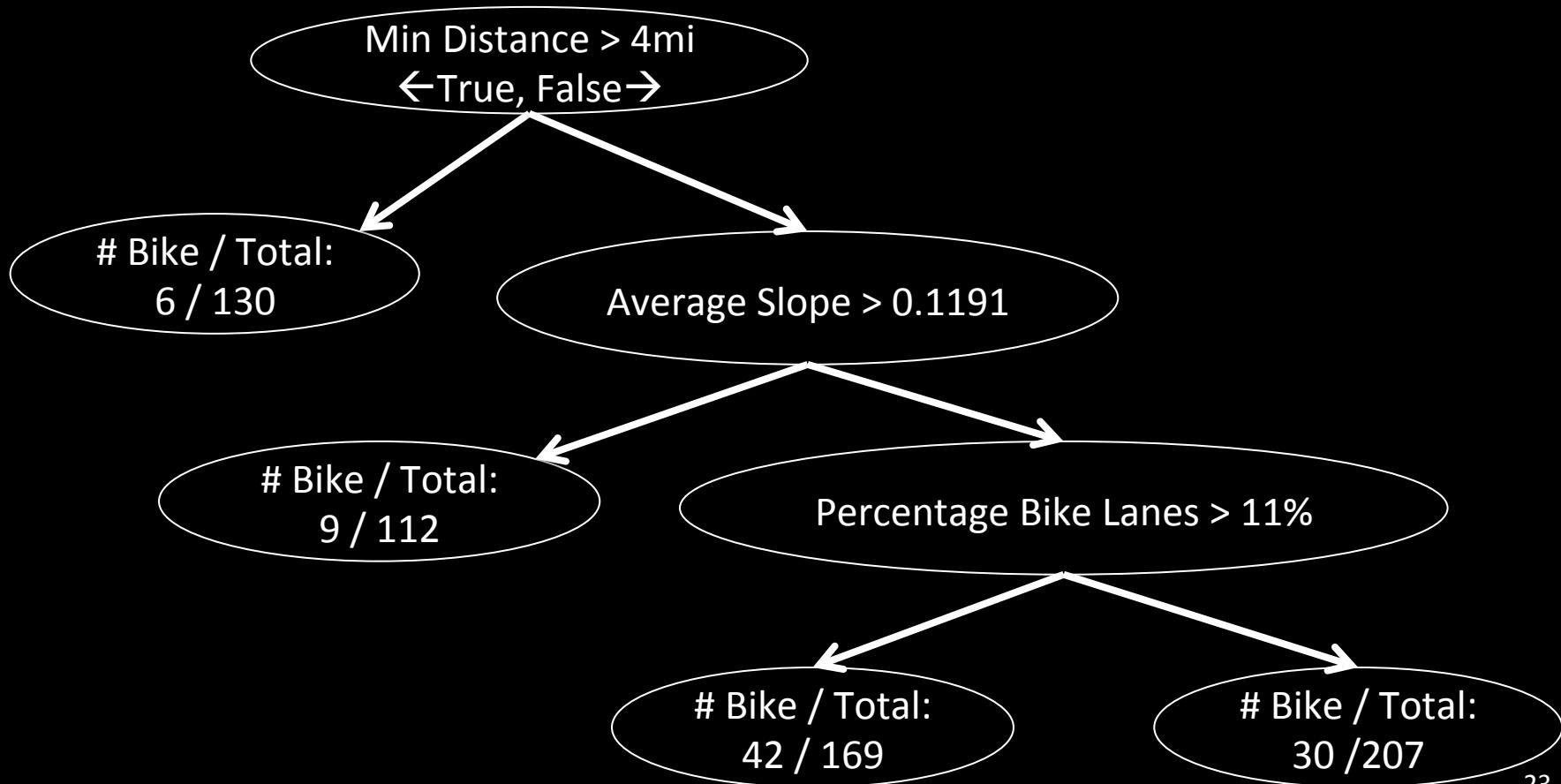
 $\neq$  $\neq$ 

# How else might we incorporate roadway-level variables?

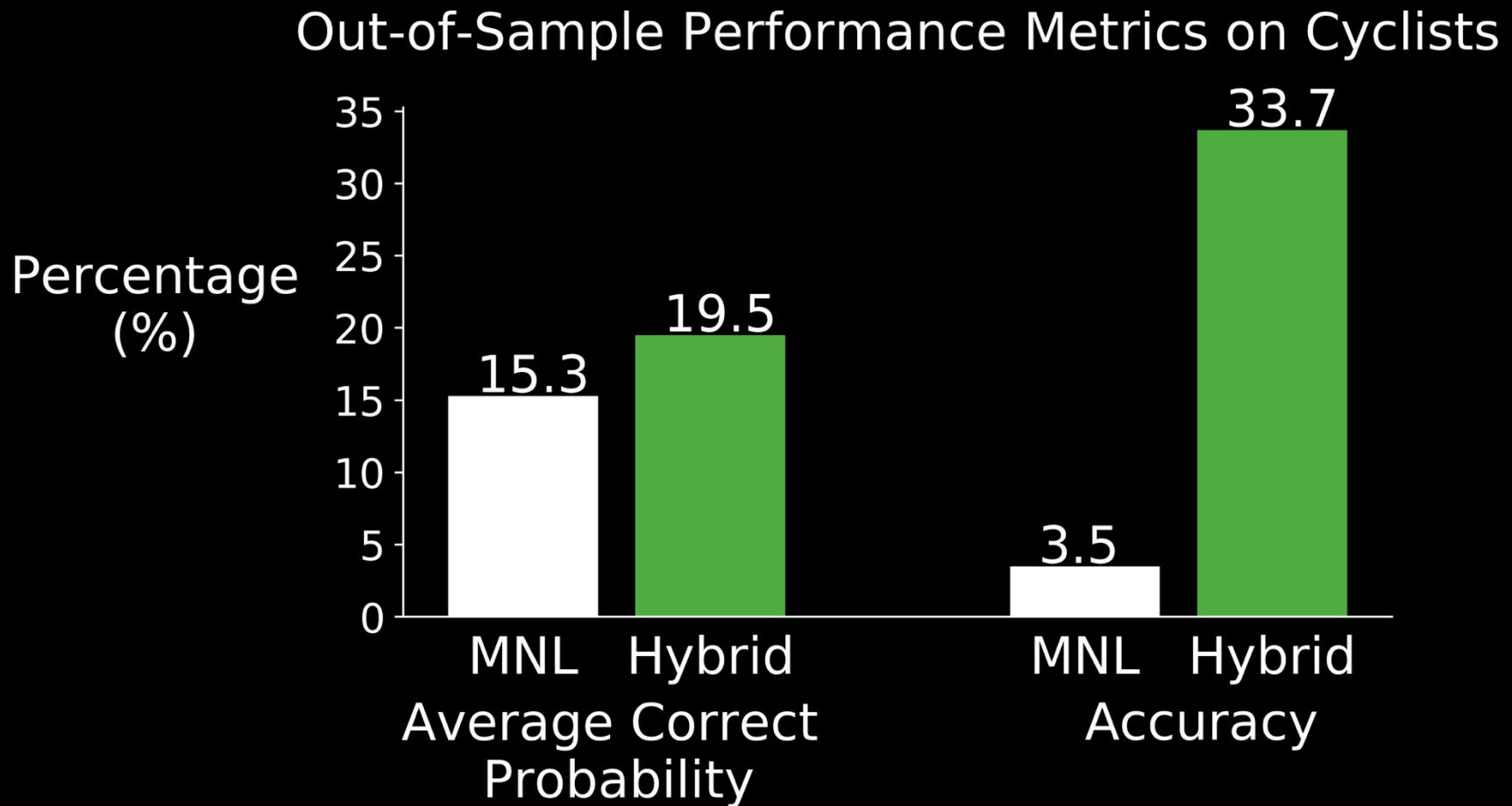


# How else might we incorporate roadway-level variables?

Decision tree for “situation description”



# Zone of Likely Travel + Decision Tree = Increased Model Accuracy



# Summary of Contributions (Part 1)

- Demonstrated feasibility and benefit of incorporating relevant roadway variables
- Increased sensitivity to roadways between origins and destinations
- Systematically and interpretably combined roadway variables

# What goes into the secret sauce?



# Behavioral Assumptions

“Homo Economicus”

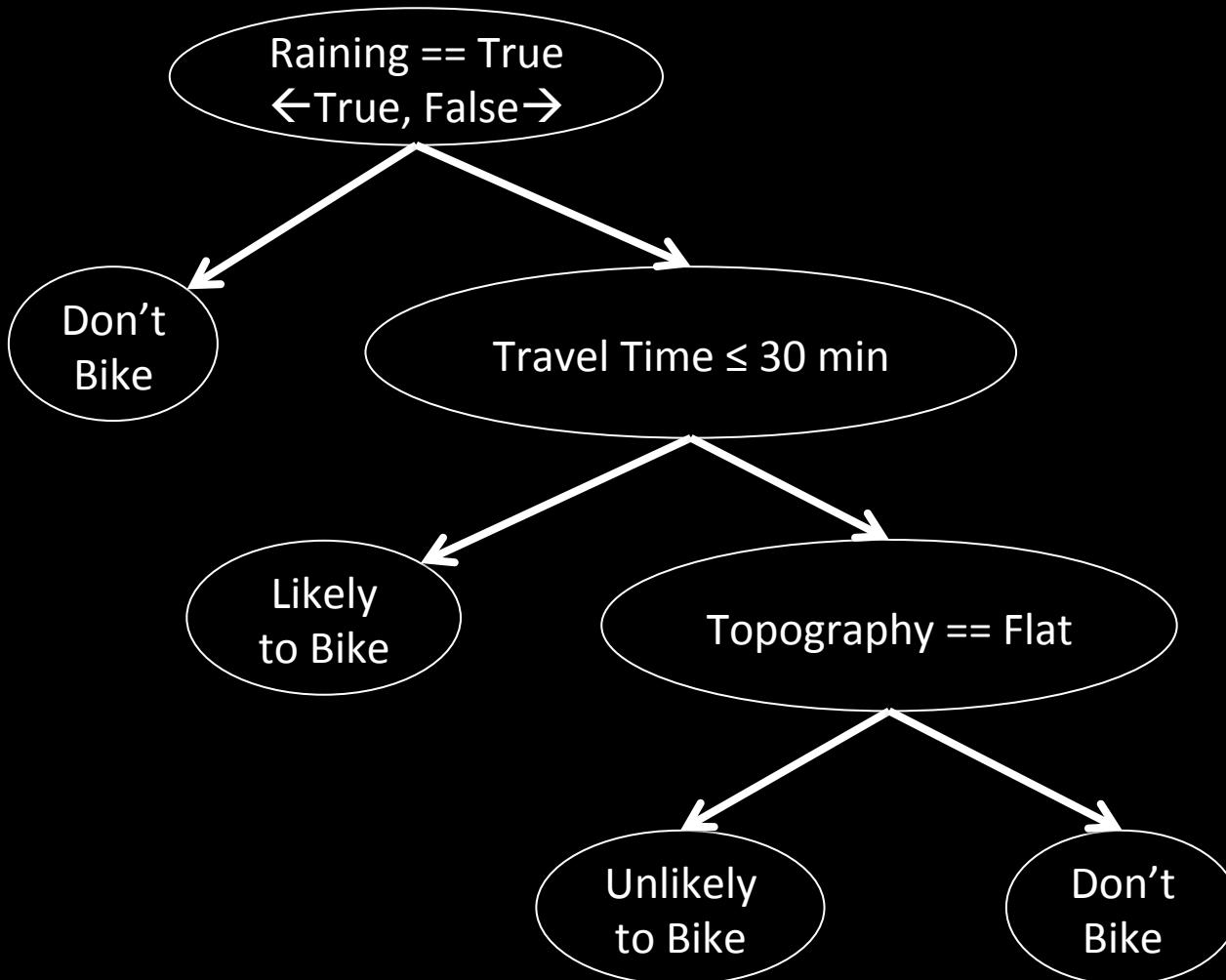


# Perfectly Rational Humans are a Myth



I WANT TO  
BELIEVE

# If-Then Decision Making



*“Pride cometh before a fall...”*

Human behavior is complex, so we must account for our uncertainty

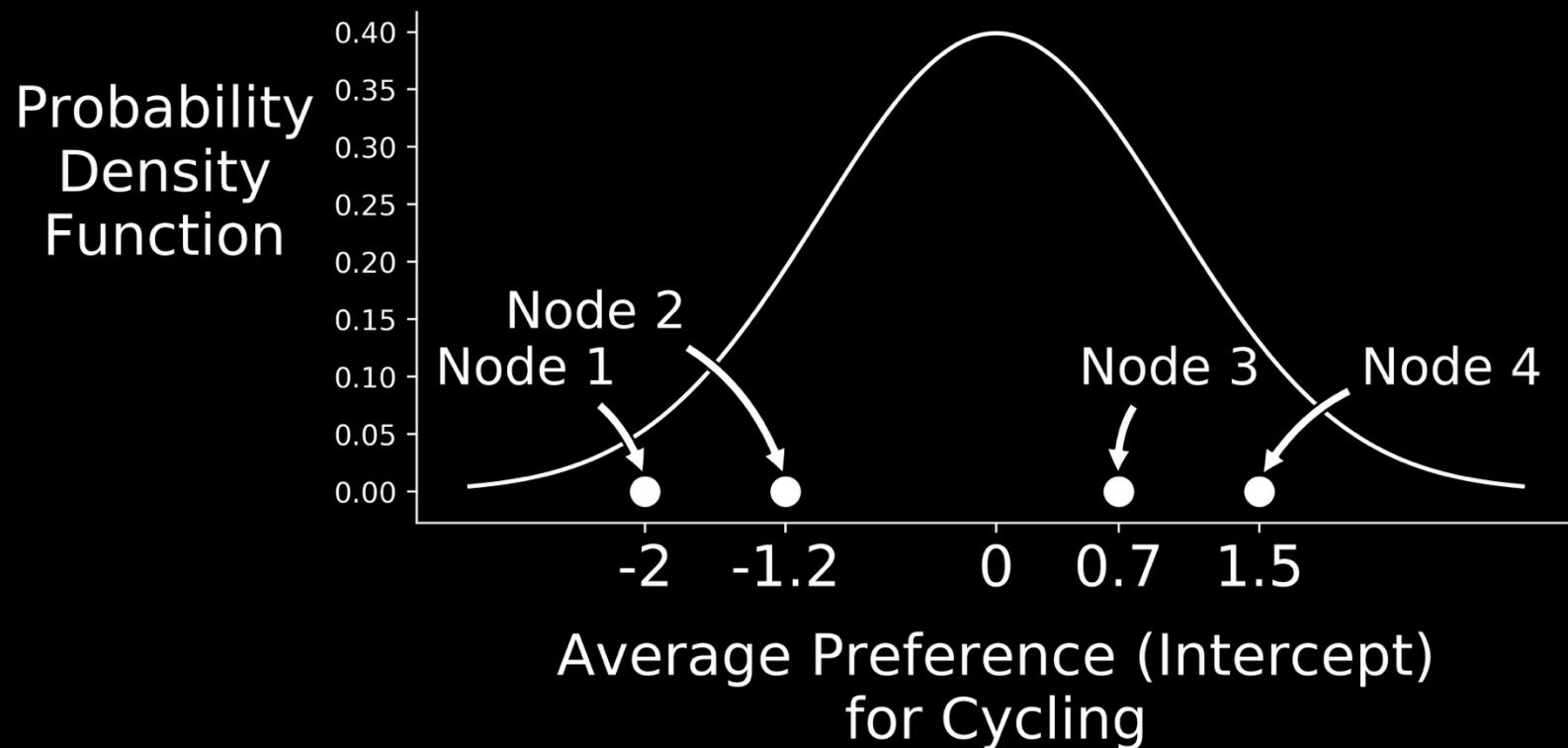
# Putting it all together: Tree-Based Gaussian Process Model

1. Many trees to describe the potential sets of if-then rules

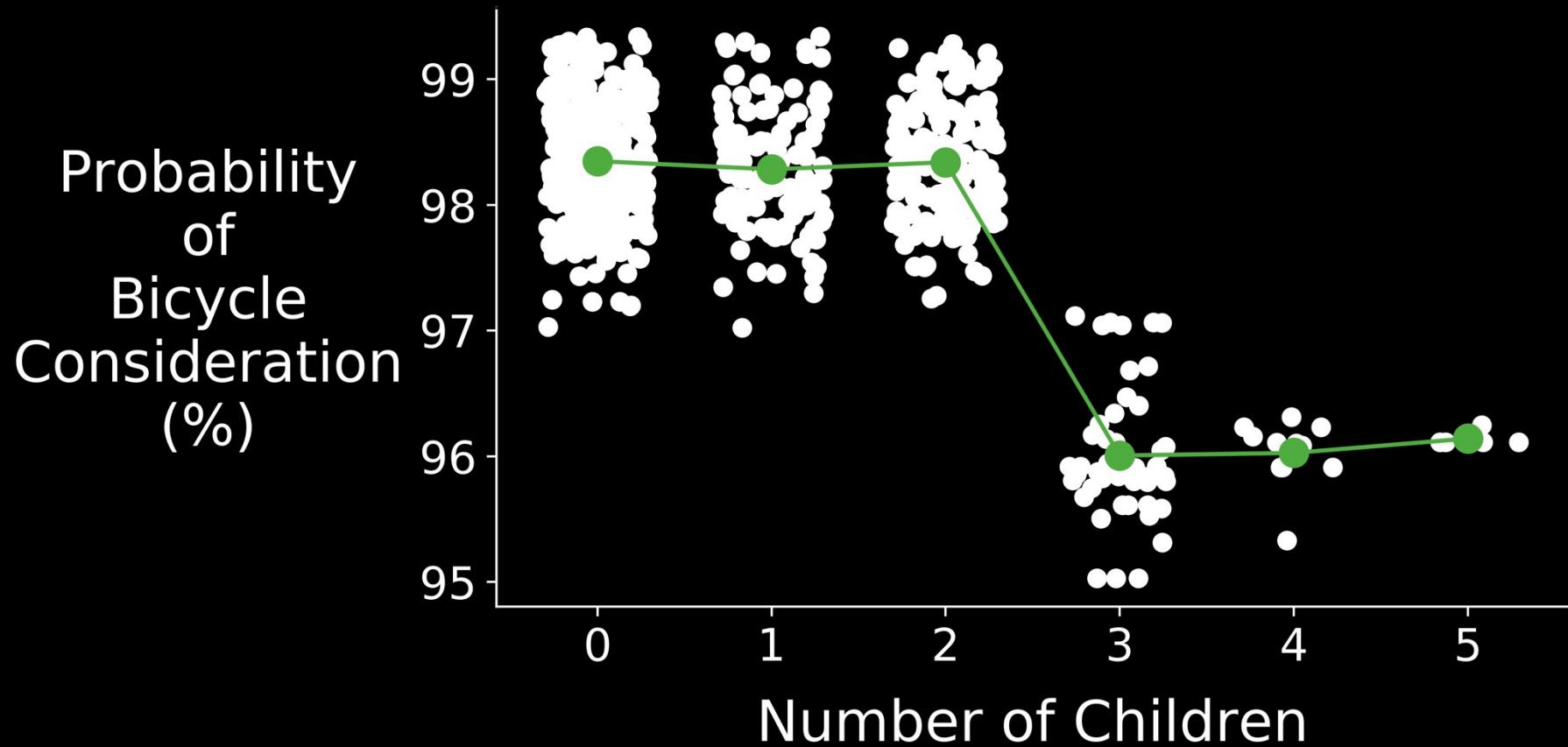


# Putting it all together: Tree-Based Gaussian Process Model

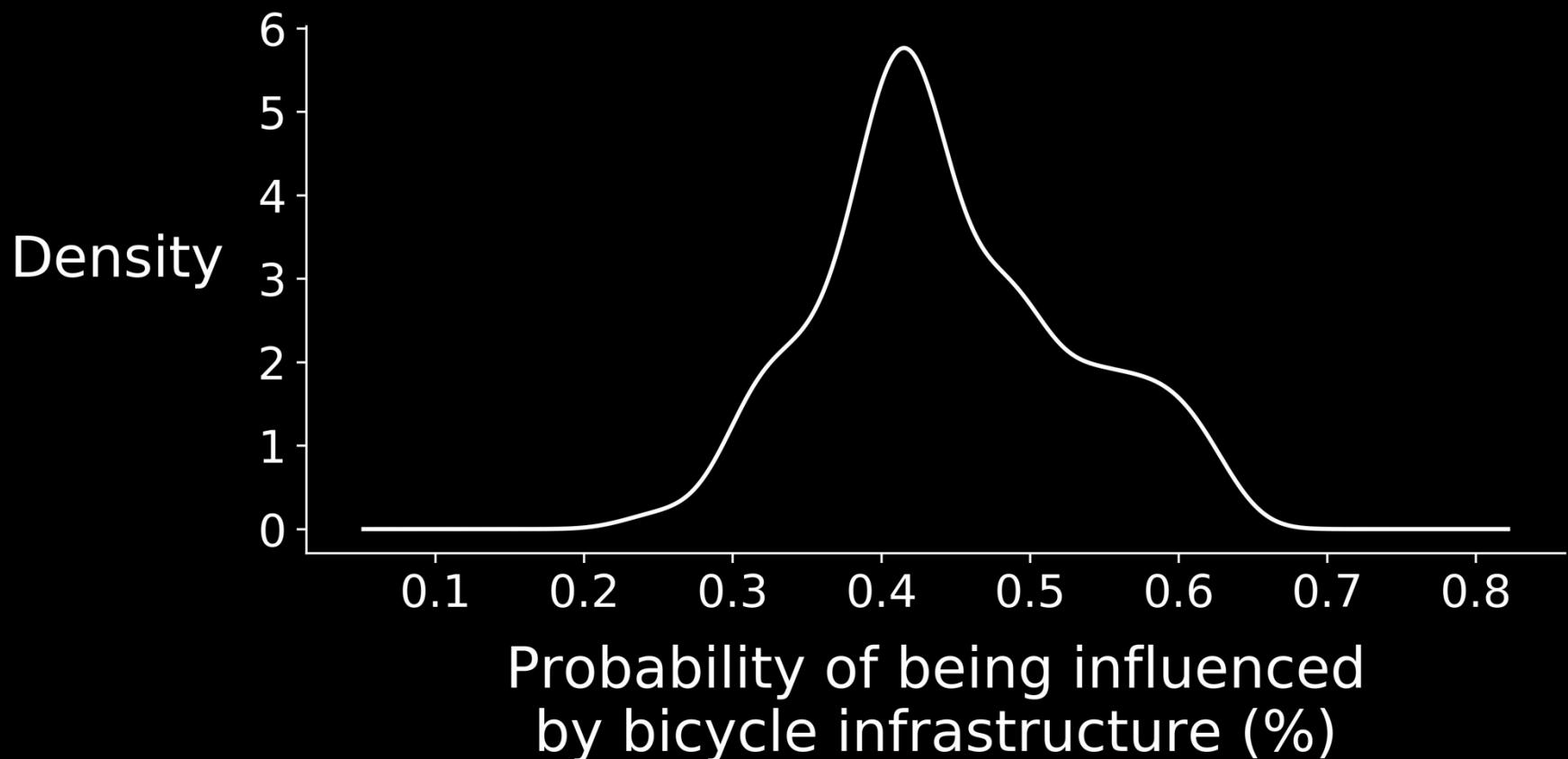
- Given a tree, the effect of each if-then rule is unique, but from a common distribution.



# What are the (preliminary) results?



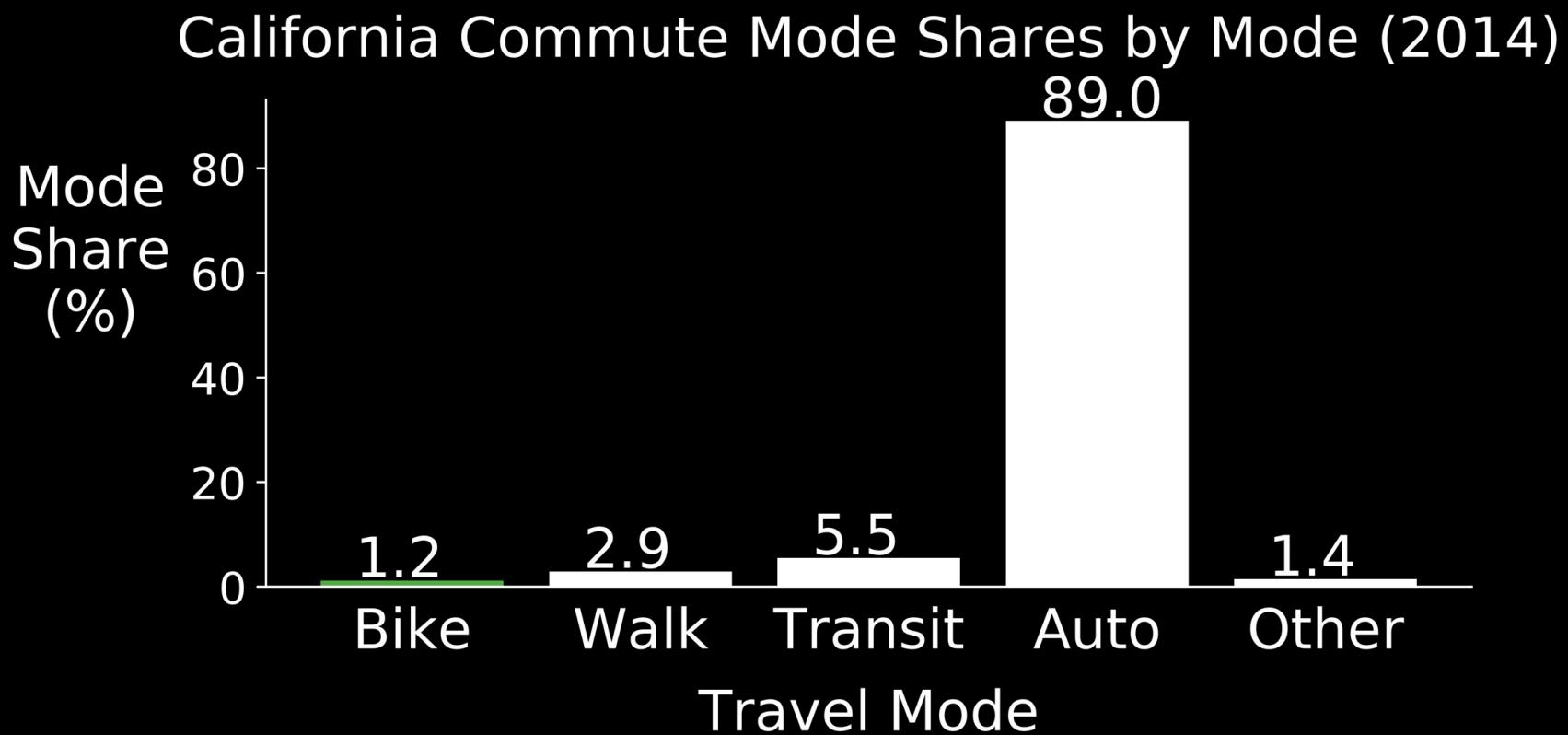
# What are the (preliminary) results?



# Summary of Contributions (Part 2):

- Flexibly account for “irrationality”
- Paints richer picture of bicycle consideration and environmental effects
- Clarifies behavioral implications of decision trees
- Econometrically links decision trees and choice models

# Cyclists are a minority



# No cyclists, No Problem?

Wrong. Low (relative) numbers of cyclists cause:

- Finite sample bias in parameter estimates

“Logistic Regression in Rare Events Data”, King and Zeng (2001)

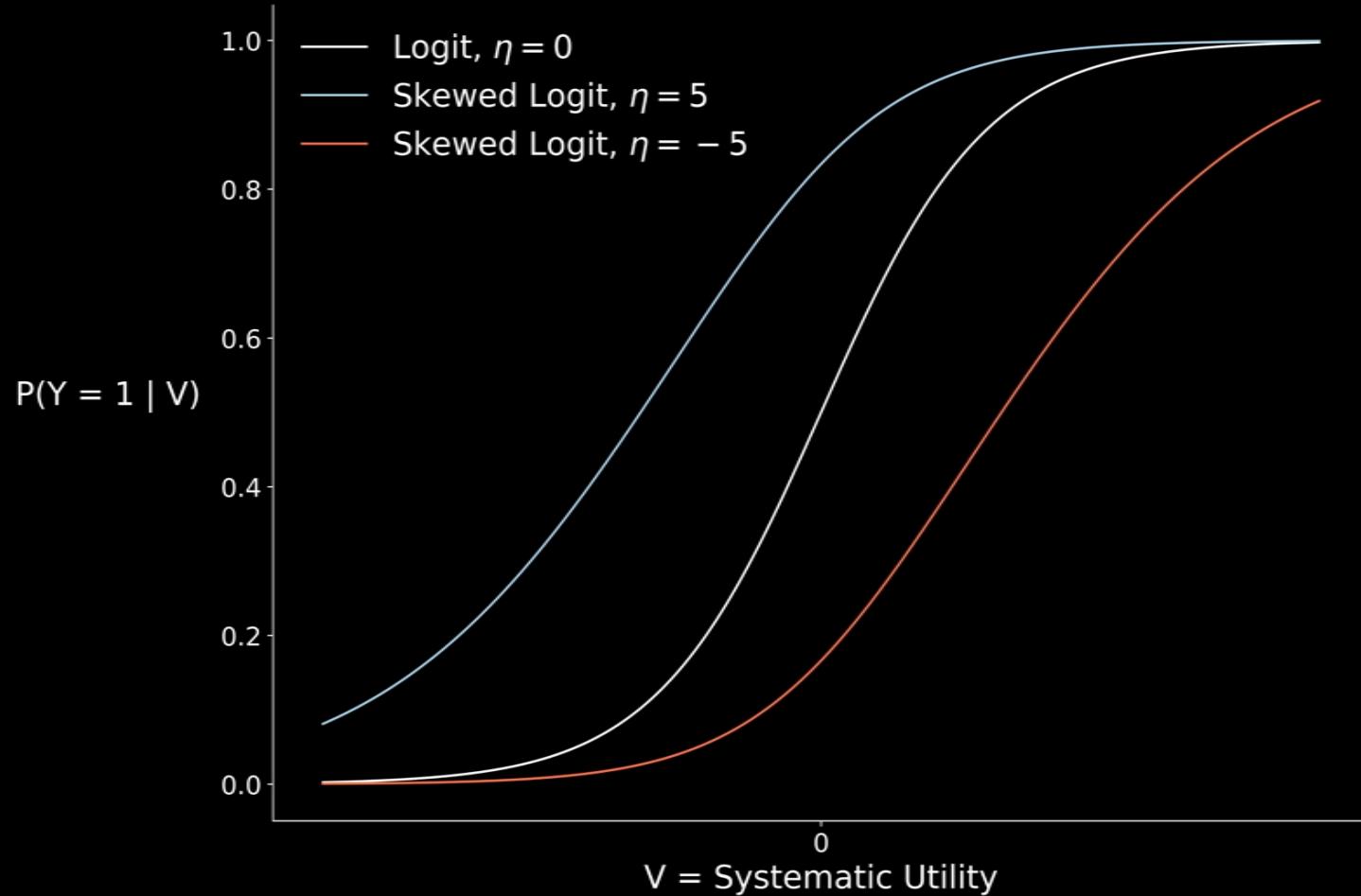
- Finite sample bias in predicted probabilities

“Logistic Regression in Rare Events Data”, King and Zeng (2001)

- Decreased predictive performance

“Predictive performance of the binary logit model in unbalanced samples”, Cramer (1999)

# How have minority alternatives been modeled in other disciplines?



# Why no Asymmetry in Transportation?

## Ad-Hoc Creation Procedures



# The Great Divide

Two Alternatives

Many Alternatives

# Summary of Contributions (Part 3)

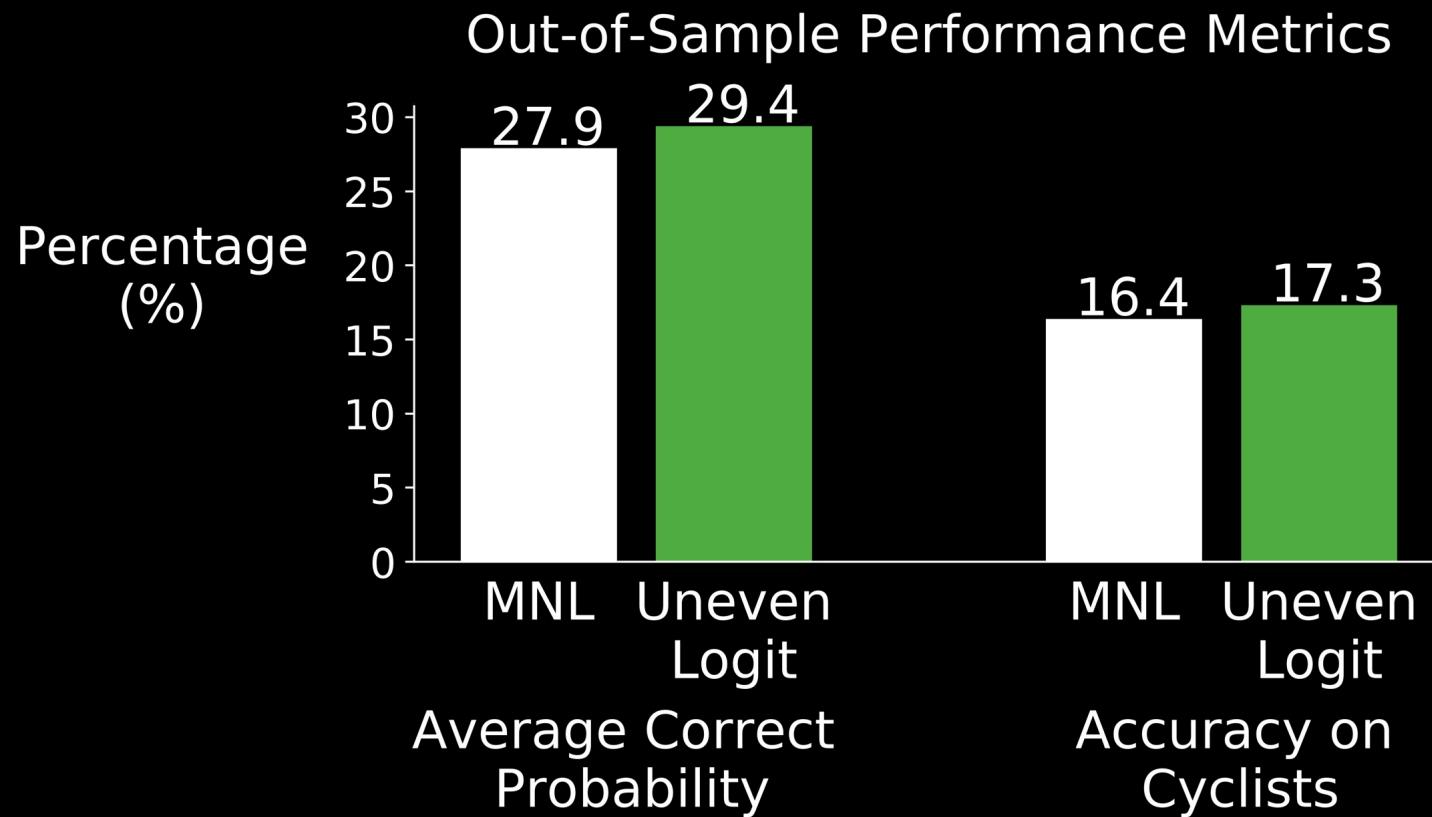
Developed algorithms to:

- create binary, asymmetric choice models
- extend binary asymmetric models to multinomial setting

Created 4 new asymmetric models as examples

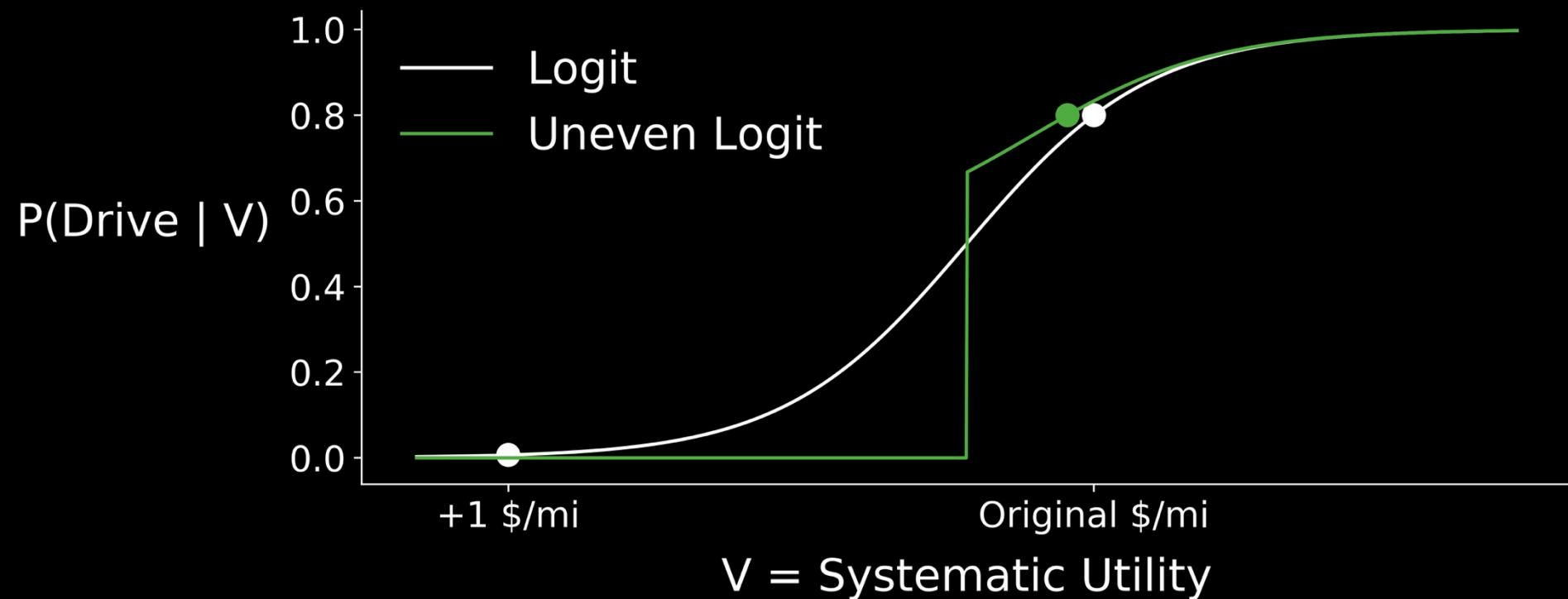
# What have we gained?

## Greater Model Accuracy



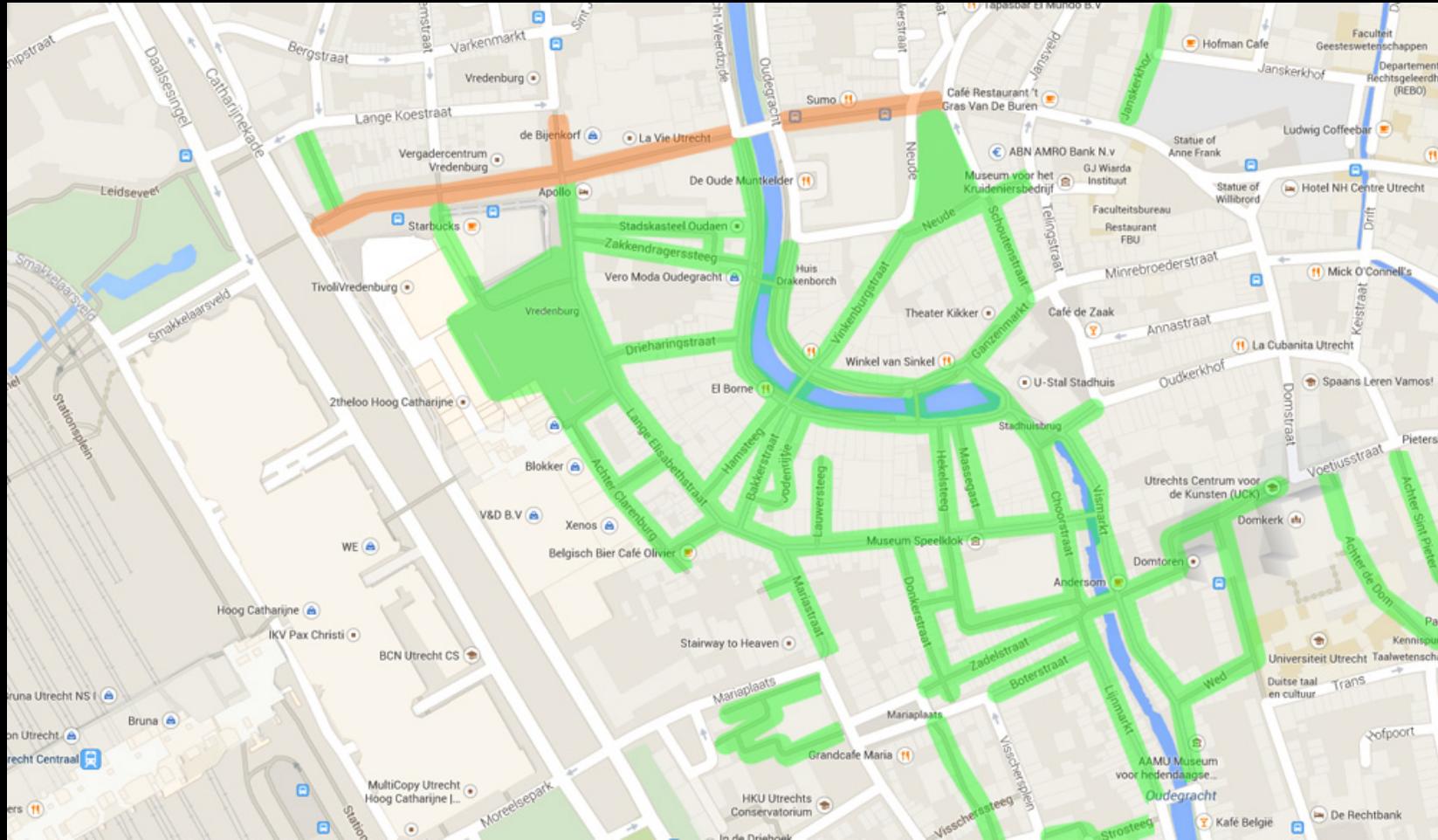
# What have we gained?

Behavioral Insights: From Adoption to Abandonment



# What have we gained?

## Stick Assessment: From Adoption to Abandonment



# Takeaways from the dissertation:

- Thoughtfully incorporate roadway-level variables
- Model the reality of ‘irrational’ travelers
- Allow for asymmetric abandonment and adoption as an explanation for low cycling numbers

# Are we good to go?

# Are we good to go?

Not Yet

# What else?

*“No Causation Without Manipulation”*

# What else?

*Optimize, optimize, optimize... proactively.*

# What else?

*Comprehensive Analysis of Heterogeneity*

Timothy Brathwaite

**THANKS FOR LISTENING!**

# ACKNOWLEDGEMENTS





















