Using machine learning to study decision making

Jeffrey R. Stevens
Department of Psychology
Center for Brain, Biology & Behavior

Which would you prefer?

$10 now

Or

$12 in one week

Intertemporal choice

Which would you prefer?

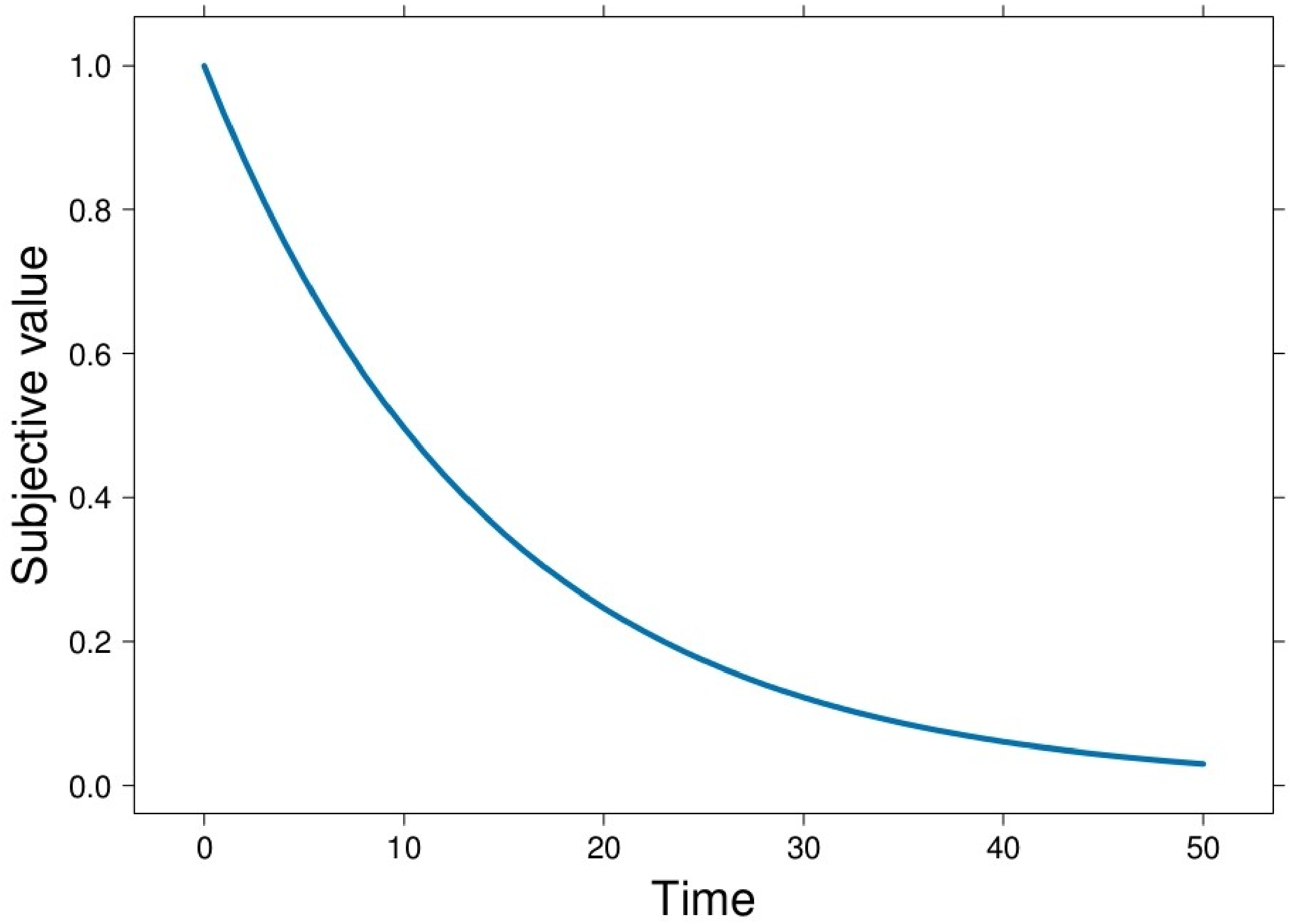
$10 now

Or

$12 in one week

Temporal discounting

Subject devaluation of the future



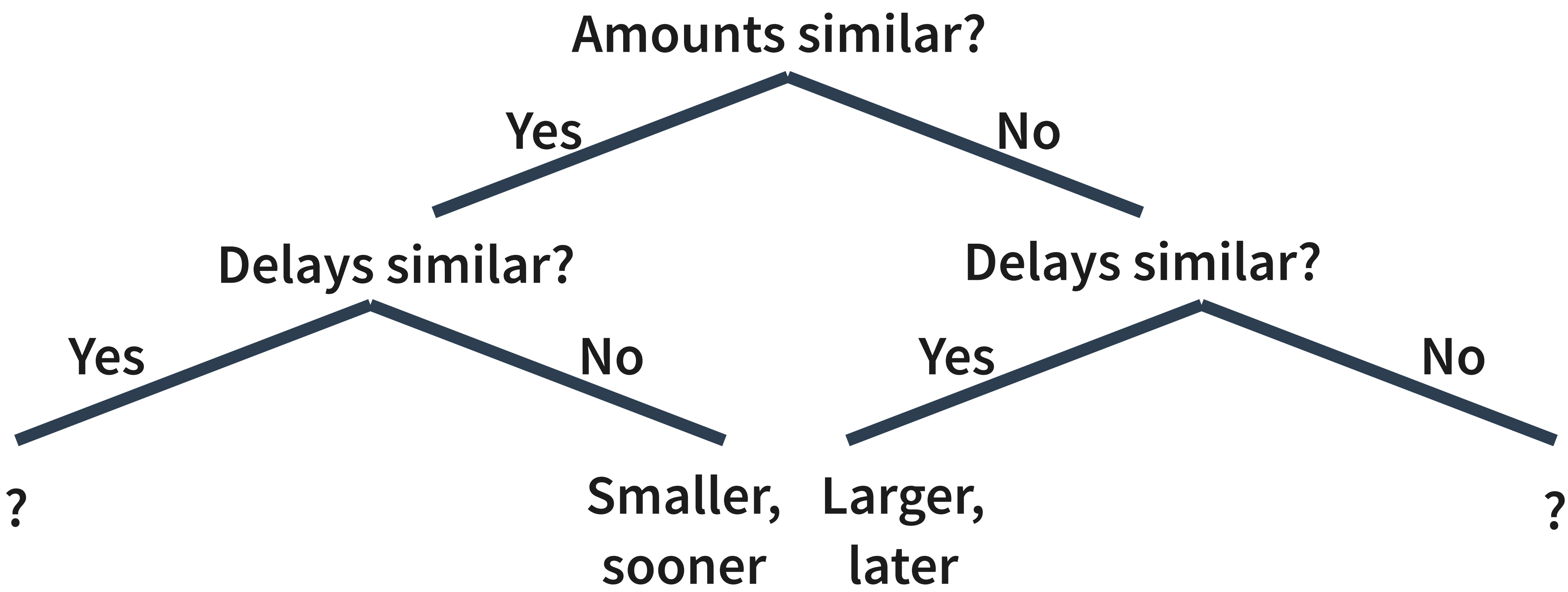
Temporal discounting

$10	now
$12	one week

Similarity

$10	now
$12	one week

Similarity


Amounts similar?
Yes
No
Delays similar?
Delays similar?
Yes
No
Yes
No
?
Smaller, sooner
Larger, later
?

Similarity vs. discounting

- Which would you prefer? $5 in 4 days or $10 in 15 days?
- Do you consider receiving $5 and $10 to be similar or dissimilar?
- Do you consider waiting 4 and 15 days to be similar or dissimilar?

Similarity vs. discounting

- Discounting predicted 72% of the intertemporal choices
- Similarity + discounting predicted 79% of the choices
- So, similarity matters

Research problem

How can we predict similarity judgments from amount and delay values?

The data

Amount and delay value pairs and similarity judgments

Two data sets:

- 65 German participants with 50 amount and 50 delay judgments
- 90 UNL participants with 43 amount and 43 delay judgments

The tools

Research questions

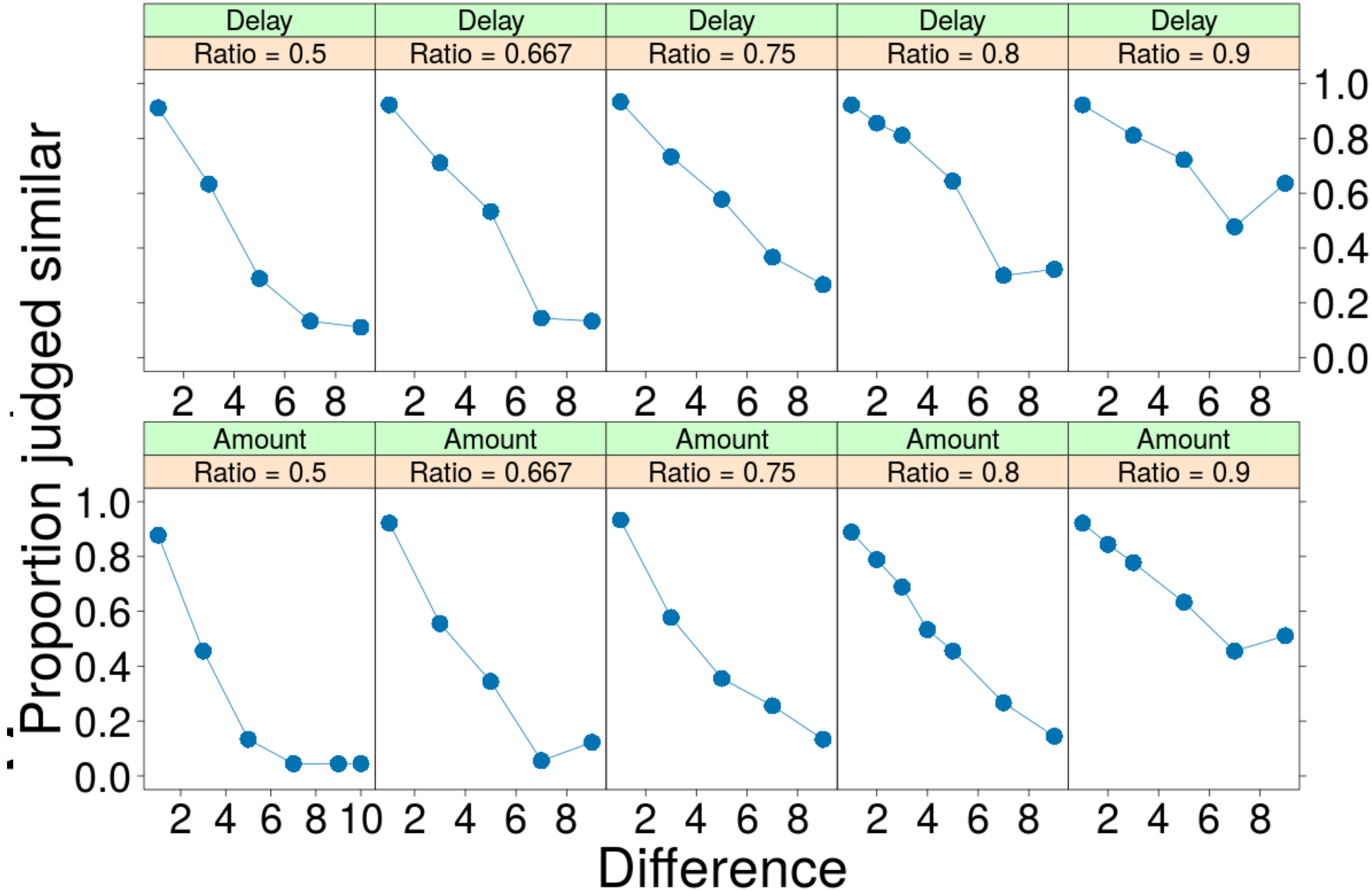
1. How do the small and large values of the reward amounts and time delays combine to predict similarity judgments?

2. Do decision trees capture similarity judgments?

How do small and large values combine to predict similarity judgments?

Predictor	Value/Function
Small value	S
Large value	L
Difference	L-S
Ratio	S/L
Mean ratio	S/[(S+L)/2]
Log ratio	log(S/L)
Relative difference	(L-S)/L
Disparity ratio	(L-S)/[(S+L)/2]
Salience	(L-S)/(S+L)
Discriminability	log(L/[L-S])
Logistic	$1/(1+e^{L-S})$

Ratio and difference



Ratio and difference

Ratio and difference independently influence judgments

Two possible reasons why:

- Ratio and difference *combine mathematically* (previous functions)
- Ratio and difference enter a *decision tree* separately in sequence

For 95-98% of participants, ratio or difference was best predictor

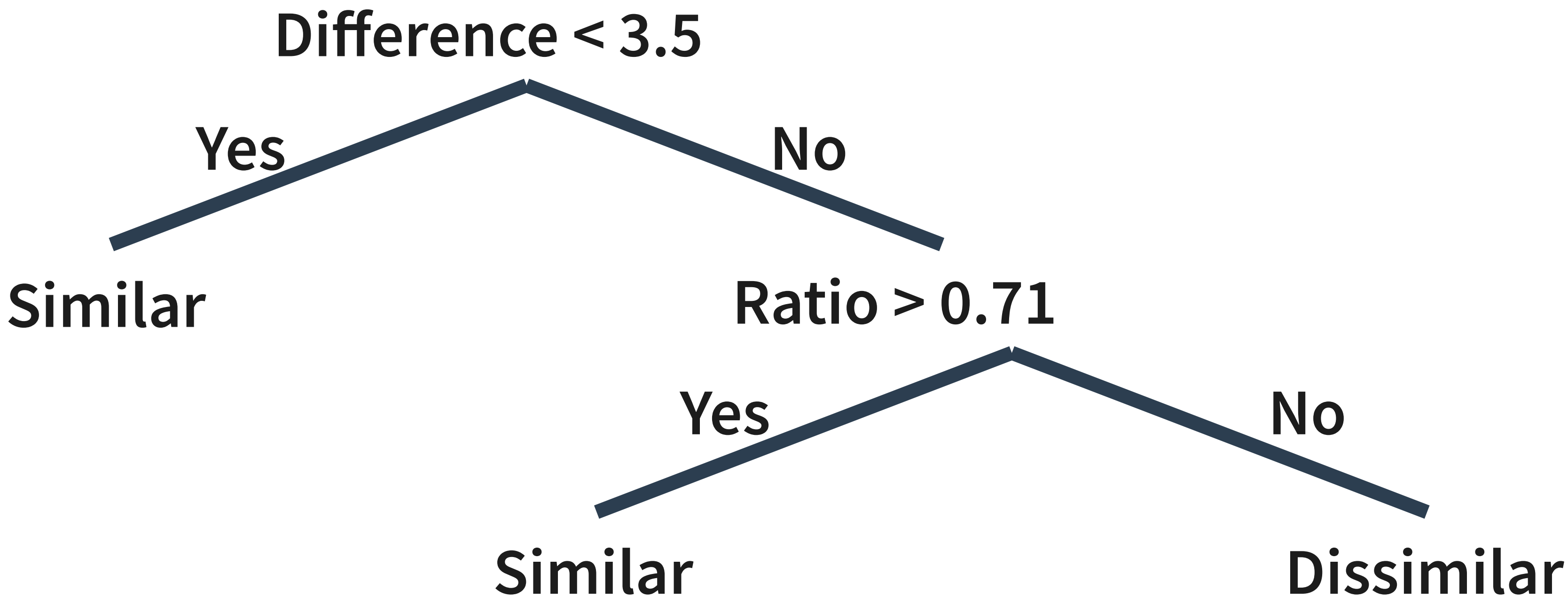
So they have separate (not combined) effects

Decision trees

Sequential decision rules for classifying outcomes based on a set of predictors

Decision trees are represented by nodes for each relevant predictor and a threshold for each predictor that divides into branches

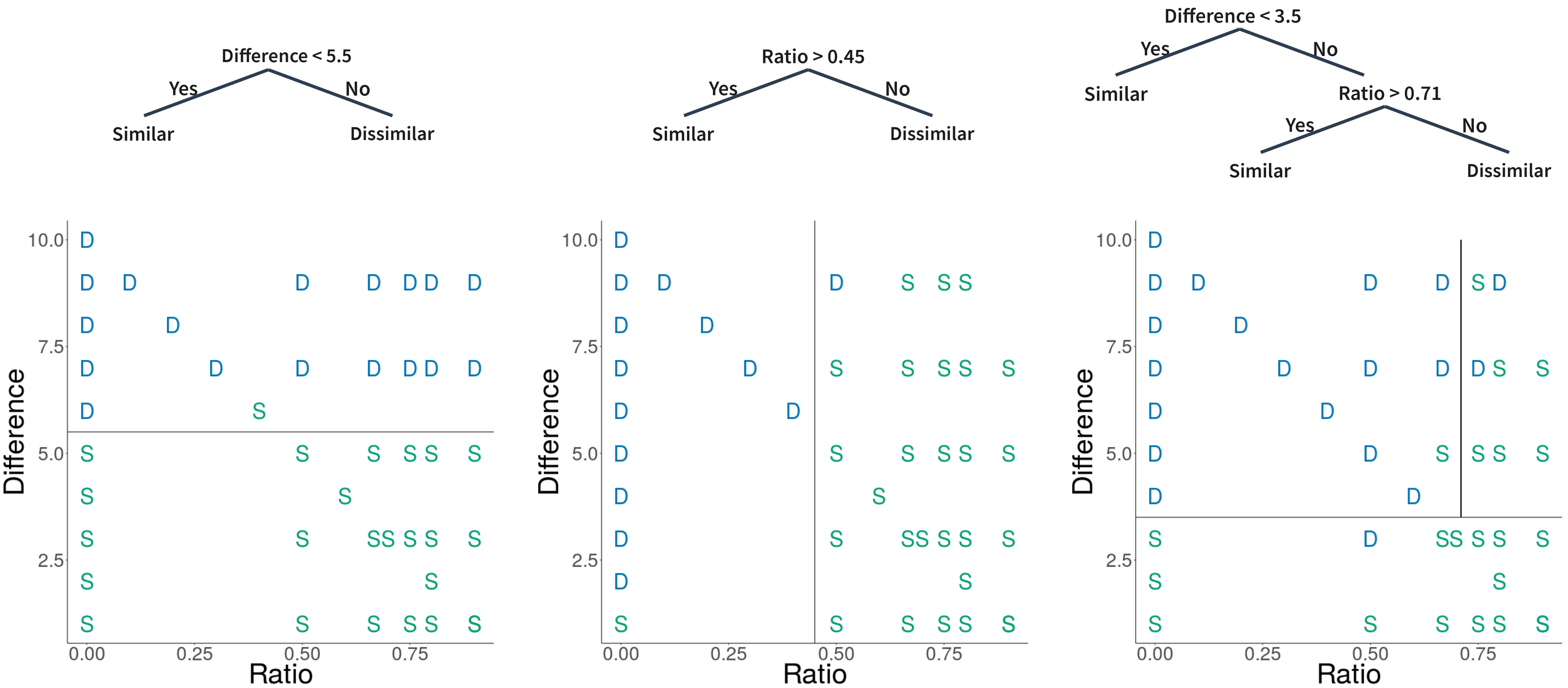
Decision trees



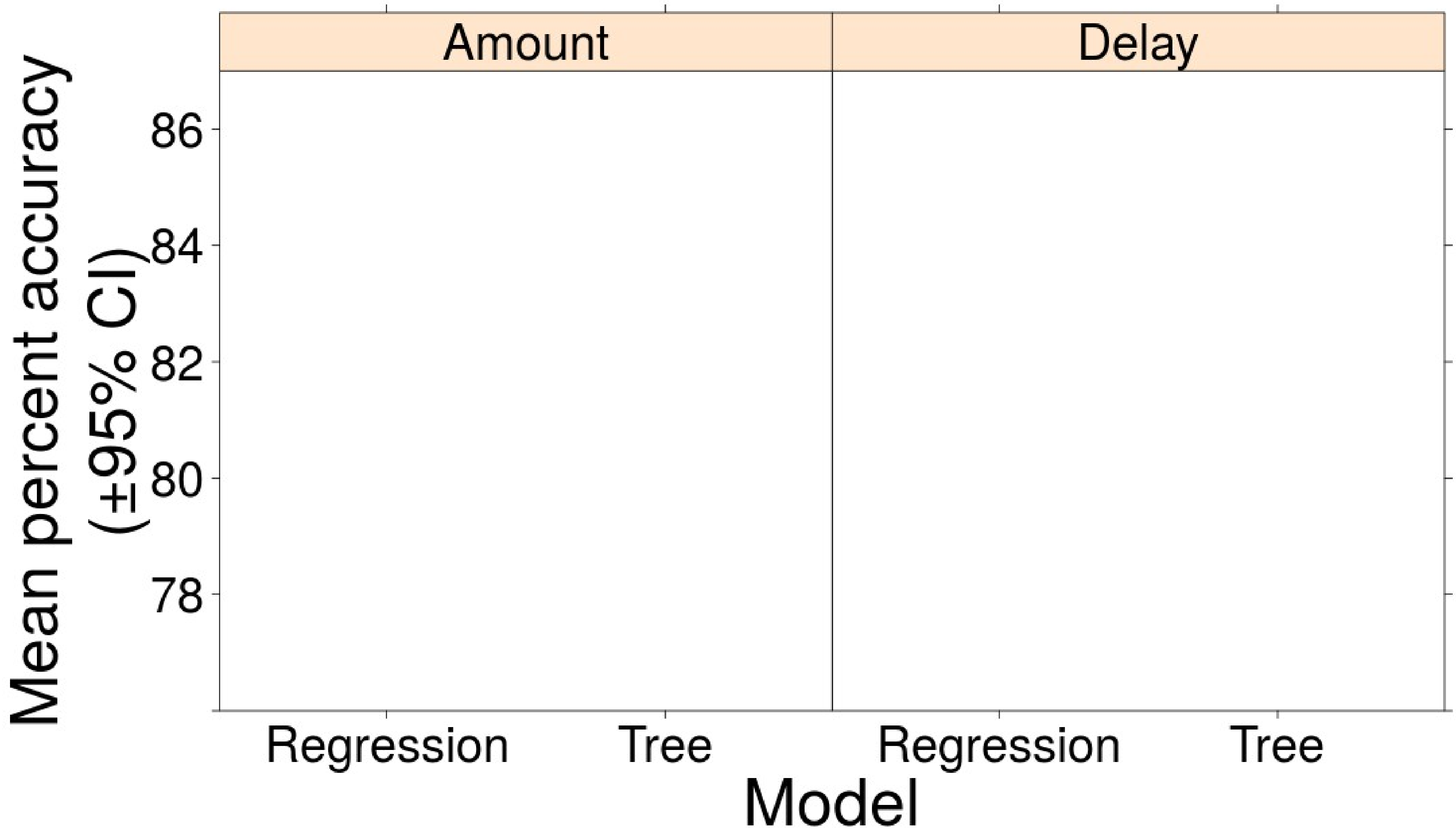
Classification and Regression Trees (CART)

- CART sequentially divides data into groups based on predictor values to most accurately classify the data according to the response variable.
- The algorithm starts with all of the data and finds the predictor and threshold value that best divides the data into two groups in a way that minimizes classification errors
- This process is then applied to each group again and continues on recursively until the last groups have no classification errors
- CART then applies cross-validation by taking a random subset of the data (training data) to create the tree then use that tree to predict the remaining test data
- Repeating this cross-validation "prunes" or removes branches that overfit the data with high cross-validated error

Decision tree



Do decision trees outpredict other models?

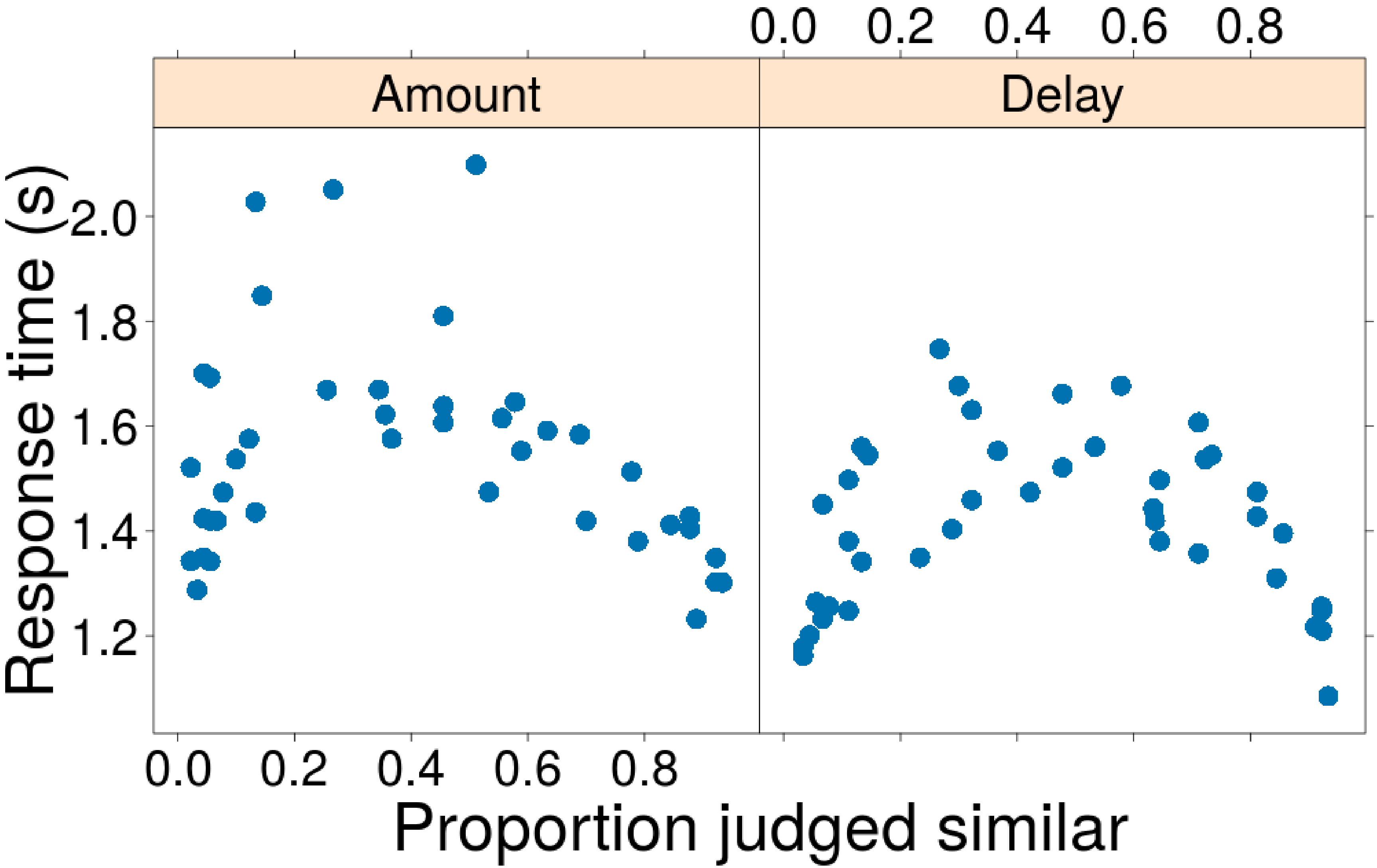


Are decision trees process models of similarity?

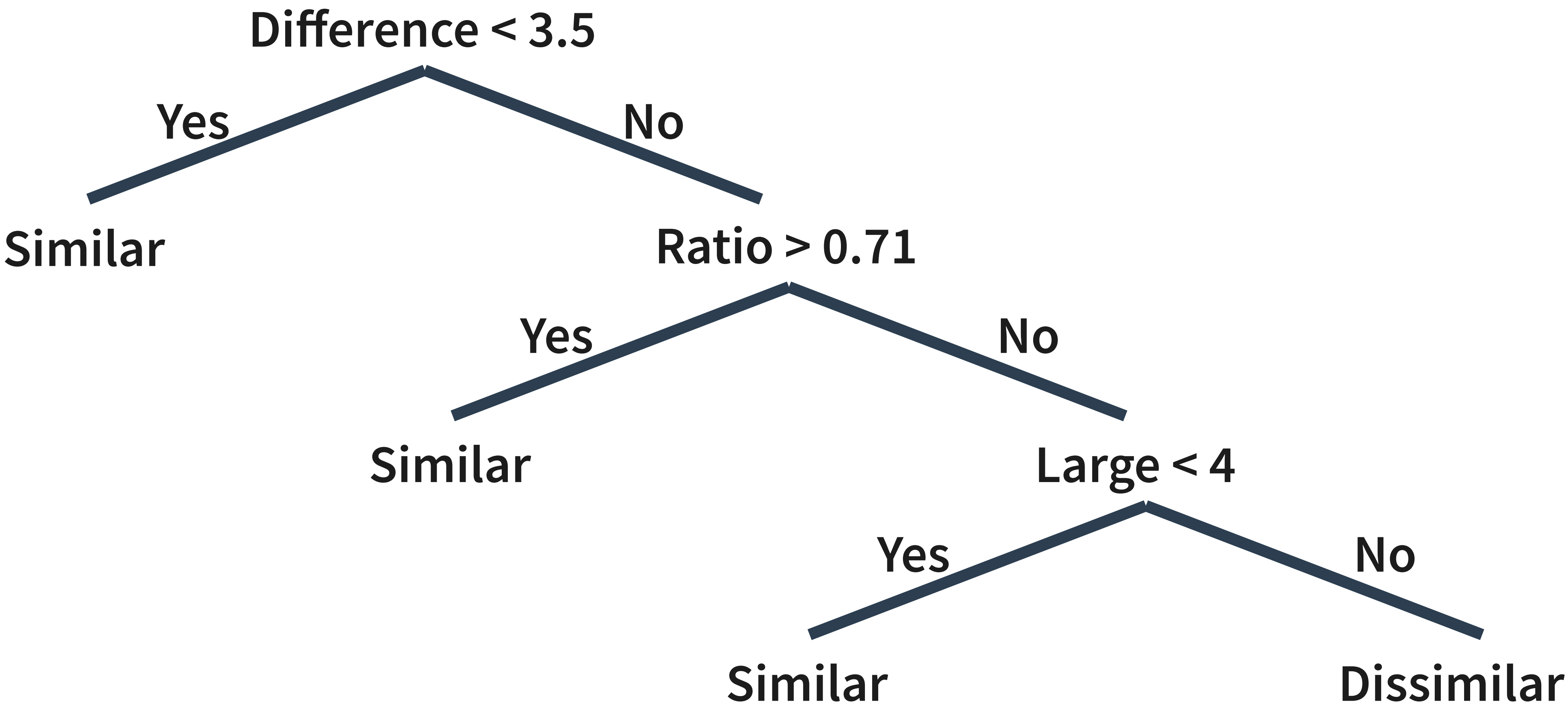
Statistical models describe a relationship between variables by fitting data to general-purpose models

Process models have theories about the process by which one variable produces another variable. For cognitive science, they capture the cognitive process generating behavior.

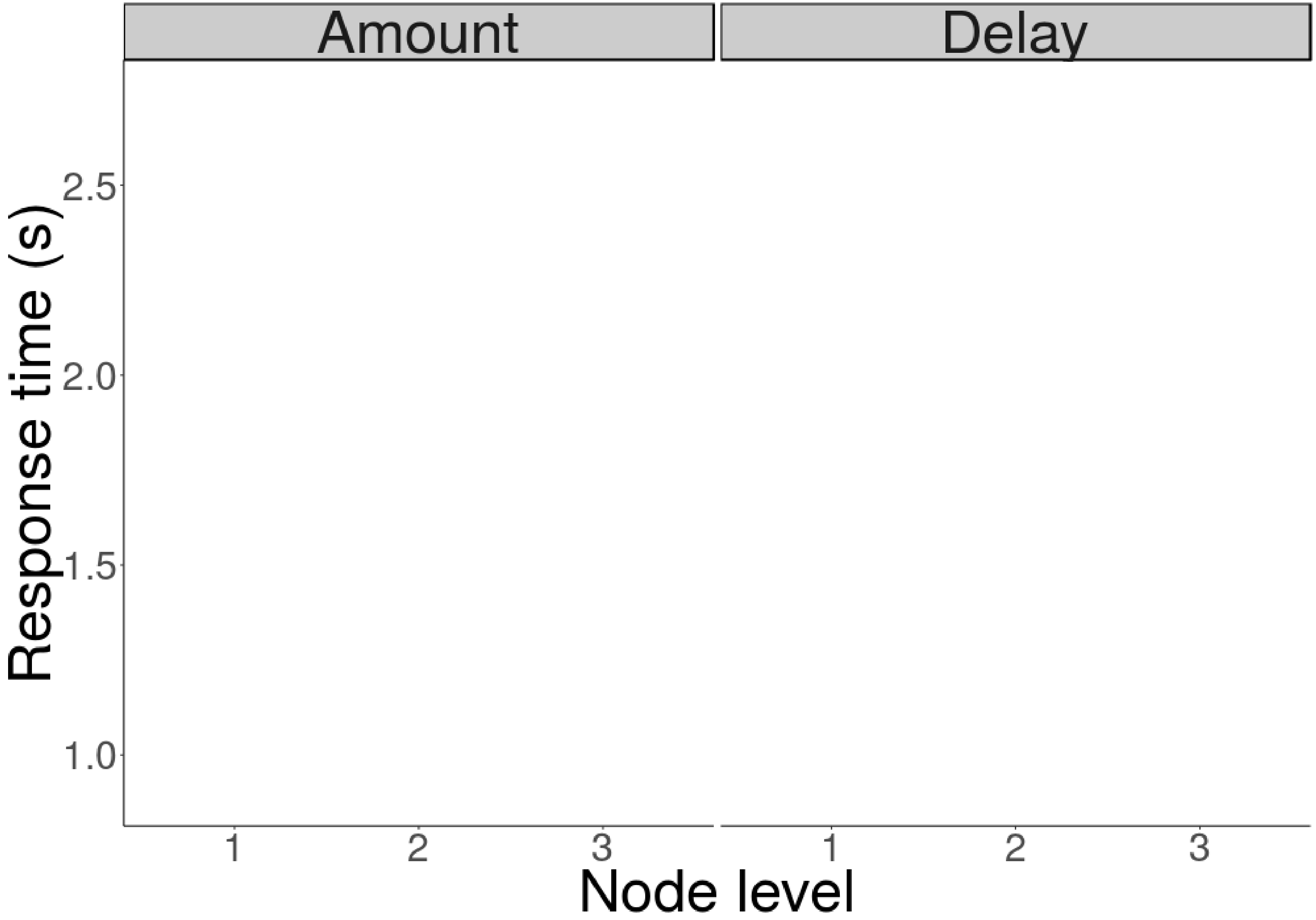
Process data



Decision tree nodes



Decision trees as process models of similarity



Research questions

1. How do the small and large values of the reward amounts and time delays combine to predict similarity judgments?

   - Ratio and difference

2. Do decision trees capture similarity judgments?

   - Yes, outcome and process data support decision trees

3. Do other algorithms do better than CART?

4. Which attributes predict best?

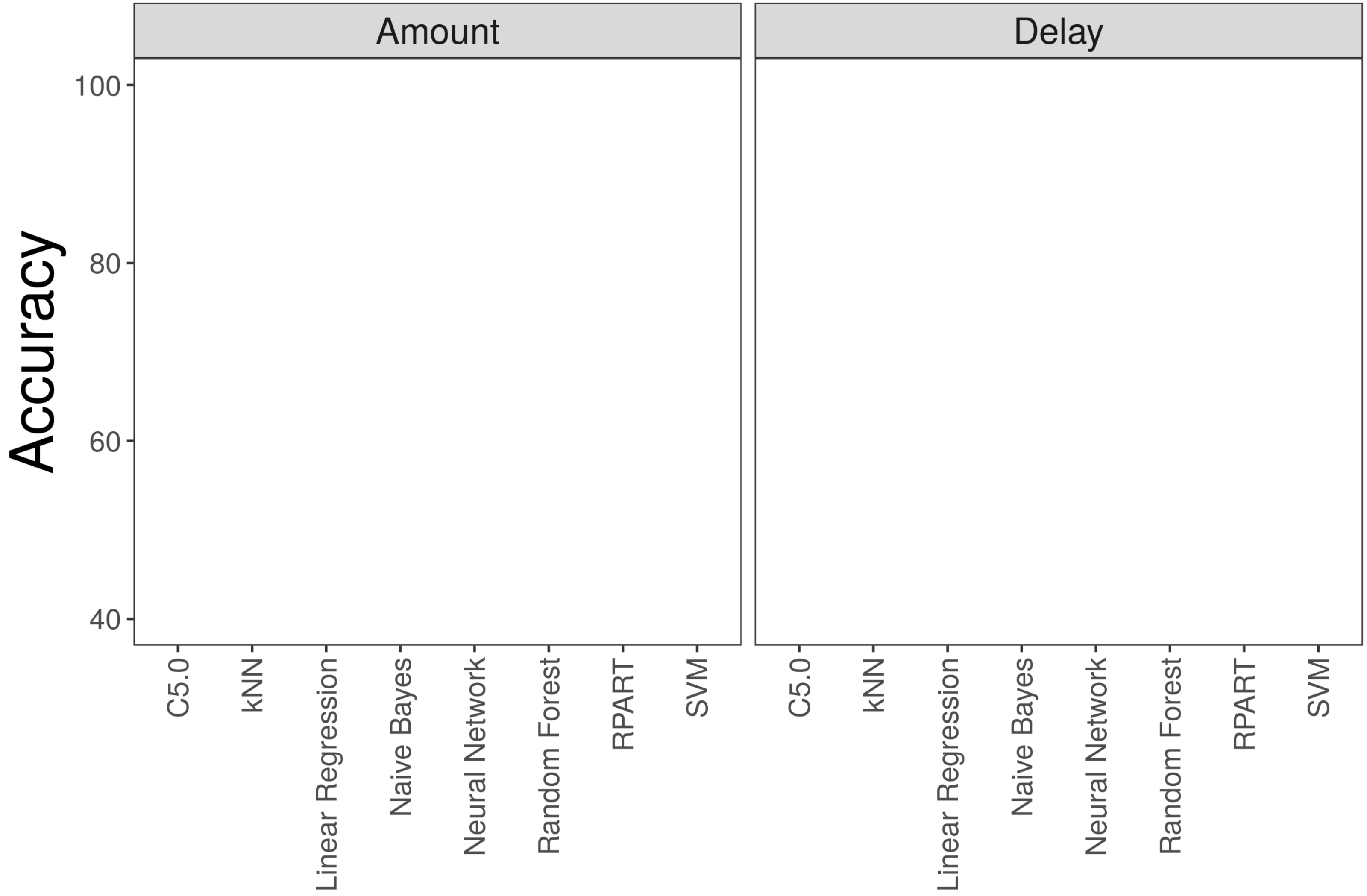
5. How does training set size influence prediction accuracy?

Best algorithm

Eight algorithms were tested based on their popularity for classification of binary data:

- CART/RPART
- C5.0
- Random Forest
- KNN
- Naïve Bayes
- Neural Network
- Support Vector Machines
- Logistic Regression

Best algorithm



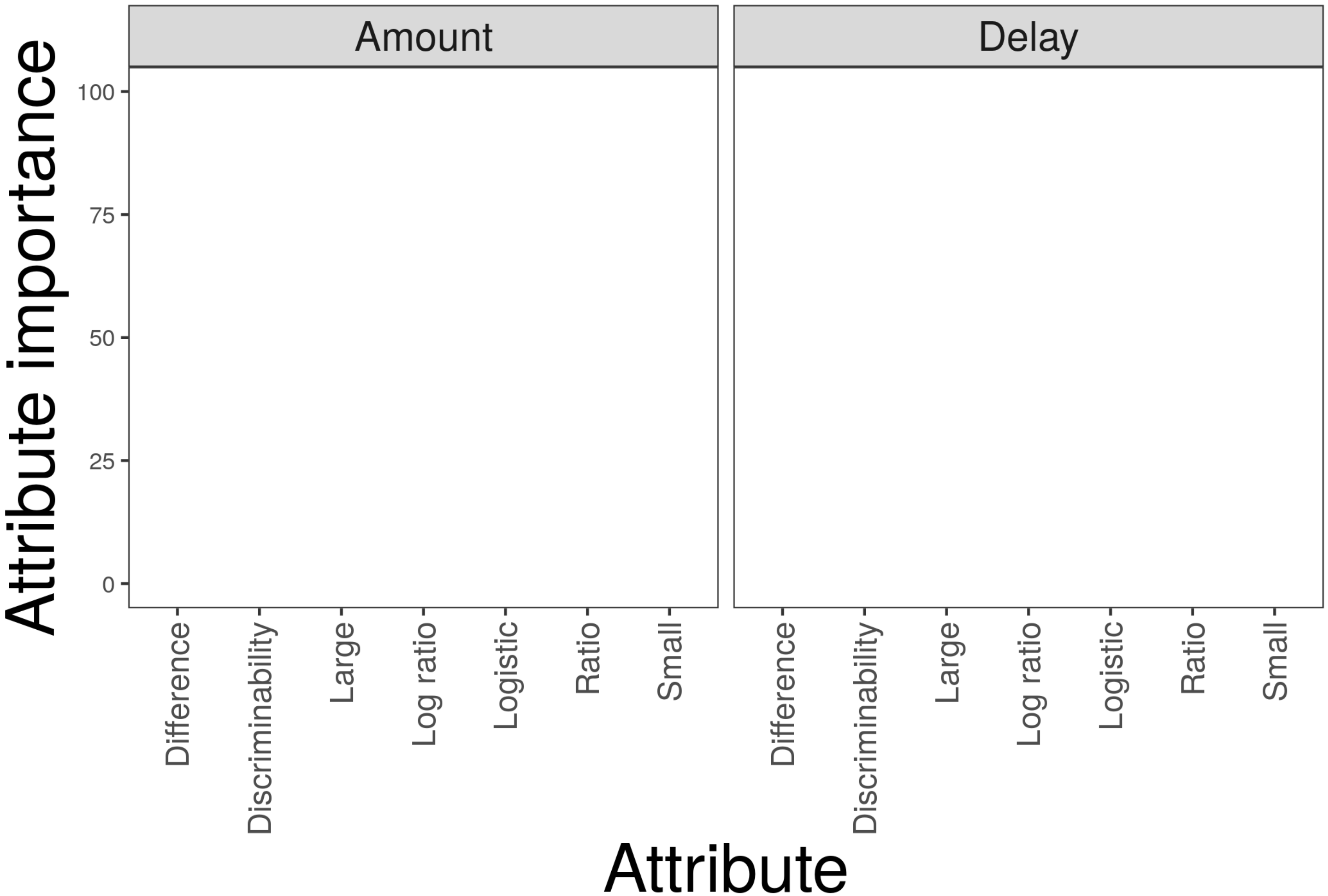
Best attributes

Removed highly correlated attributes

- Small
- Large
- Difference
- Ratio
- Log ratio
- Discriminability
- Logistic

Examined attribute importance across all algorithms

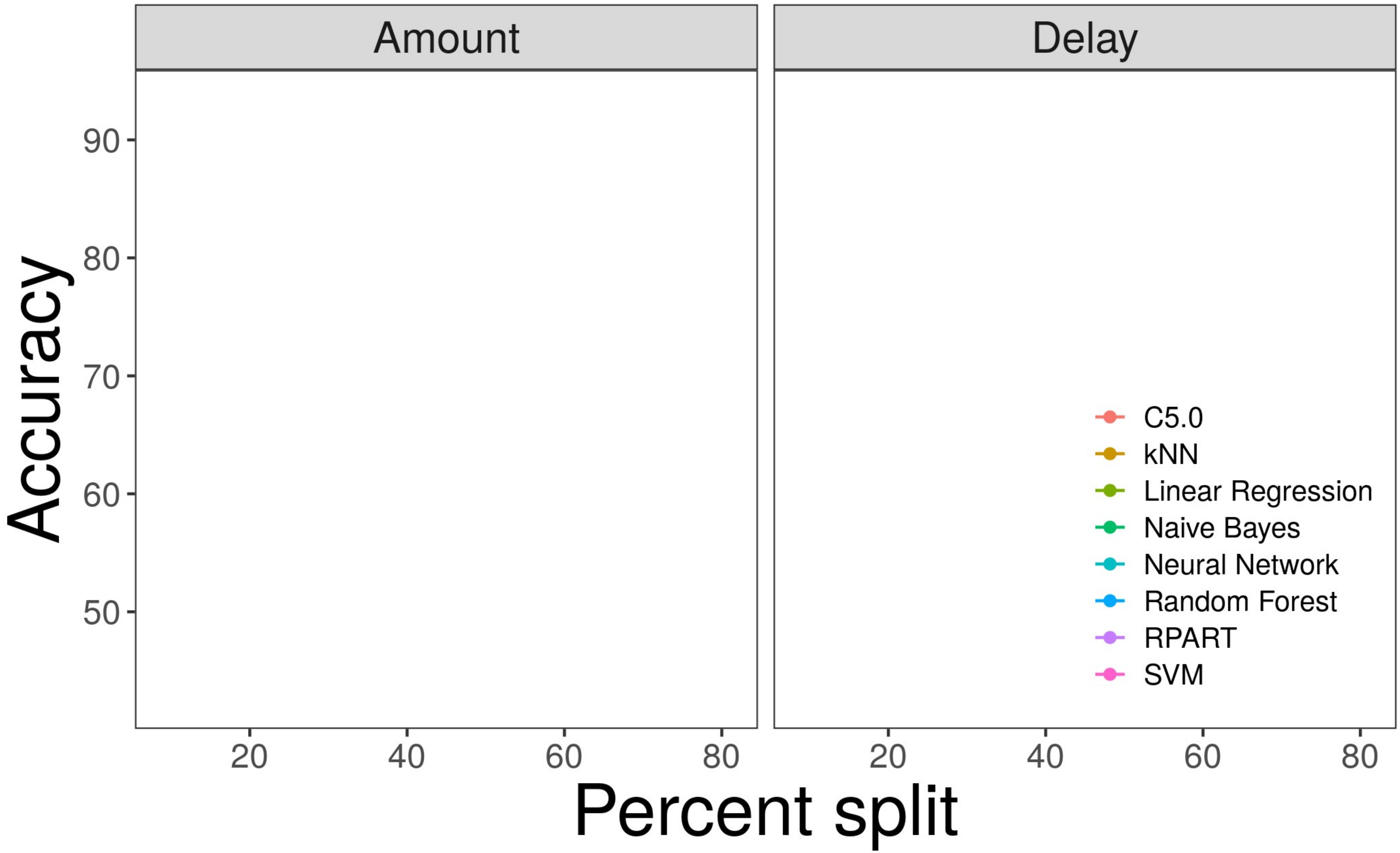
Best attribute



Training set size

How few questions can we ask to be able to predict similarity judgments?

Training set size



Research questions

1. How do the small and large values of the reward amounts and time delays combine to predict similarity judgments?

2. Do decision trees capture similarity judgments?

3. Do other algorithms do better than CART?

   - Yes, but we don't understand them

4. Which attributes predict best?

   - Ratio and difference, maybe logistic and discriminability

5. How does training set size influence prediction accuracy?

   - Depends on algorithm, but many work with low sample size

Take home message

- Understanding similarity judgments can provide insight into intertemporal, risky, and strategic choice
- Which may inspire interventions to foster better decisions
- Machine learning provides novel and useful ways to understand decision-making data