Supplementary Materials: Improving measurements of similarity judgments with machine-learning algorithms

Jeffrey R. Stevens[1], Alexis Polzkill Saltzman[1], Tanner Rasmussen[1], & Leen-Kiat Soh[1]

[1] University of Nebraska-Lincoln

Table S1
Predictors

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

Note. Table from Stevens & Soh (2018).

Correspondence concerning this article should be addressed to Jeffrey R. Stevens, B83 East Stadium, Center for Brain, Biology & Behavior, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA. E-mail: jeffrey.r.stevens@gmail.com

Table S2
Predictor importance

Algorithm	Predictor importance calculation
C5.0	Percentage of training set samples that fall into all the terminal nodes after the split
CART	Reduction in the loss function (e.g., mean squared error) attributed to each variable at each split
kNN	Area under the ROC curve
Naive Bayes	Area under the ROC curve
Neural network	Absolute values of node weights
Random forest	Difference between prediction accuracy on the out-of-bag portion of the data and after permuting each predictor variable, averaged over all trees and normalized by the standard error
Regression	Absolute value of the t–statistic for each model parameter

Note. Drawn from *caret* package documentation (Kuhn, 2020).

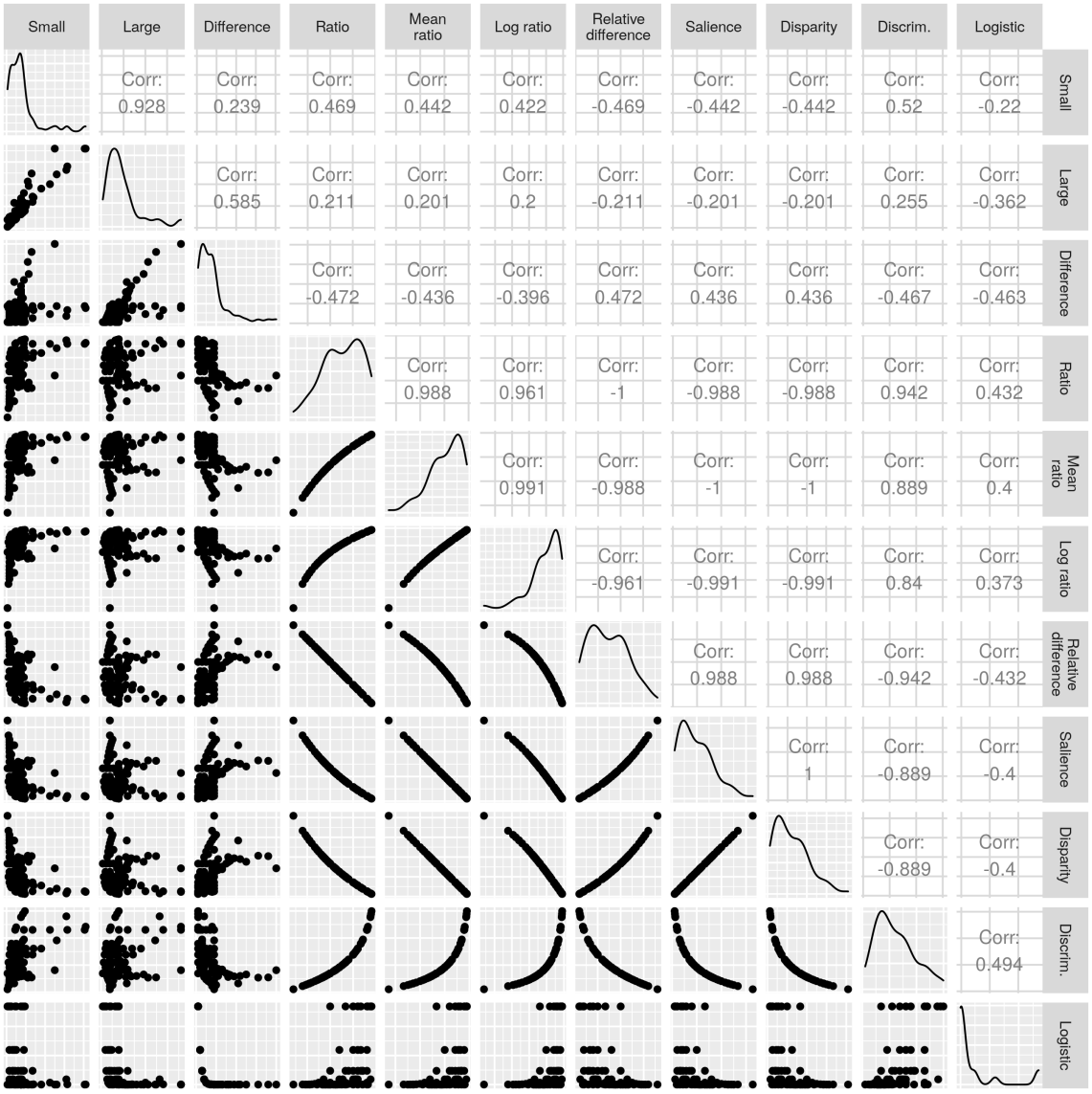

Figure S1. Pairwise correlations for *amount* similarity judgment predictors. Diagonal shows histogram, below diagonal shows correlation plots, above diagonal shows correlation coefficients. Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.

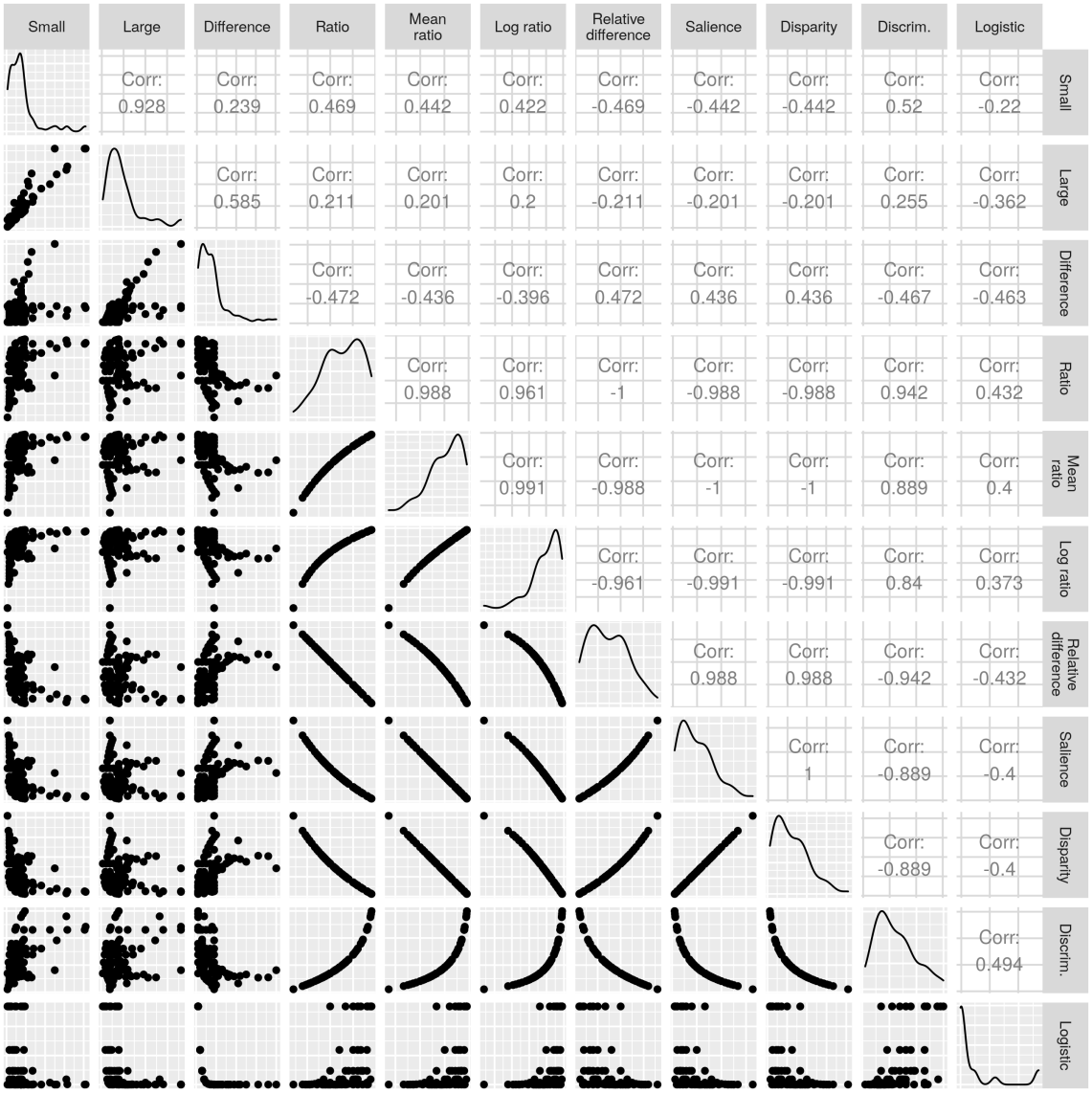

Figure S2. Pairwise correlations for *delay* similarity judgment predictors. Diagonal shows histogram, below diagonal shows correlation plots, above diagonal shows correlation coefficients. Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.

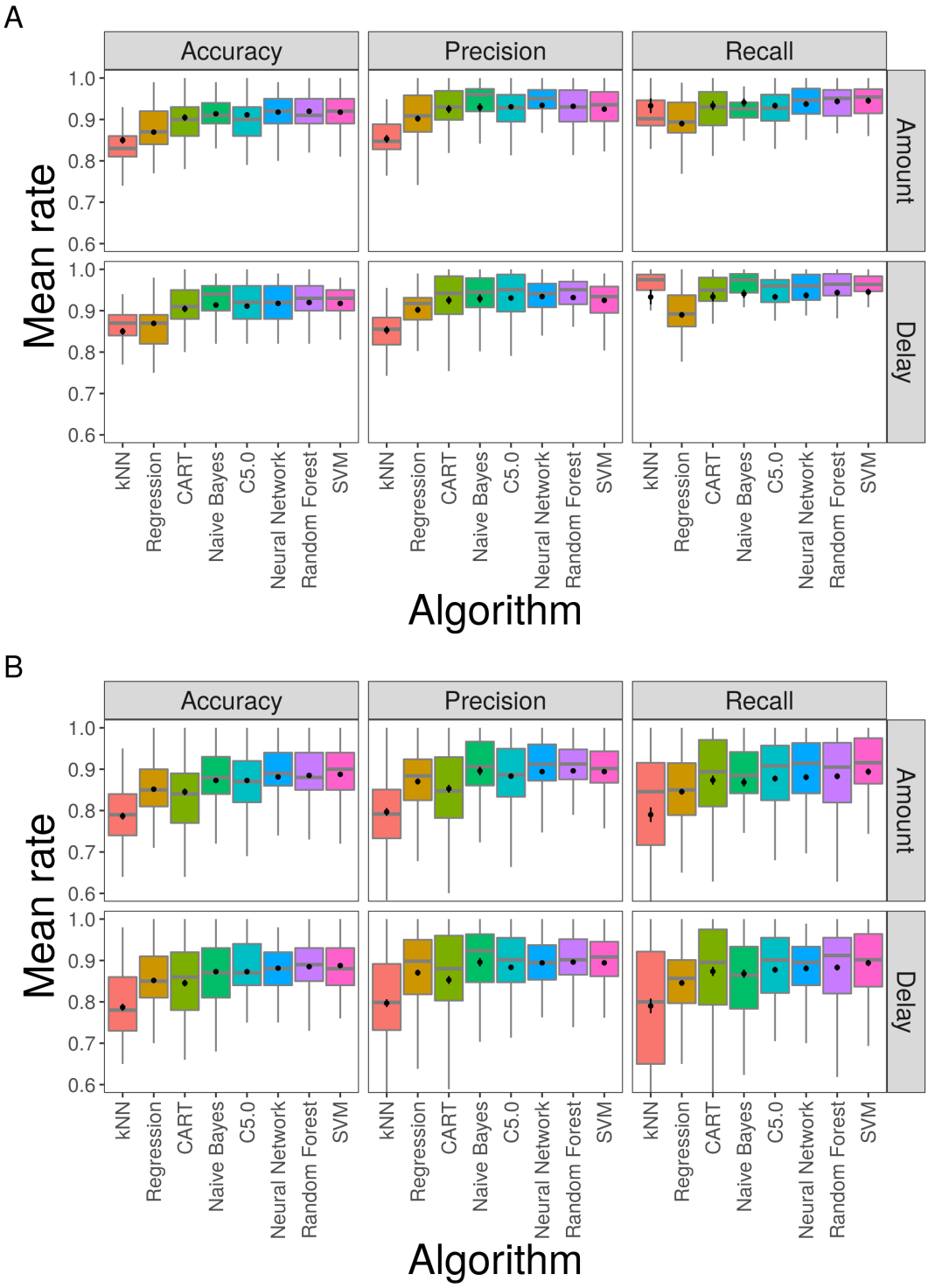

Figure S3. Accuracy, precision, and recall rates for each algorithm, judgment type, and data set (A = Data set 1, B = Data set 2). Algorithms are ordered by overall accuracy rates. Dots represent means, error bars represent within-subjects 95% confidence intervals, boxplot horizontal lines represent medians, boxes represent interquartile range, whiskers represent 1.5 × interquartile range. Outliers are not shown. Note the y-axis is truncated at 0.6 to enlarge the presentation of the means and confidence intervals. Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.

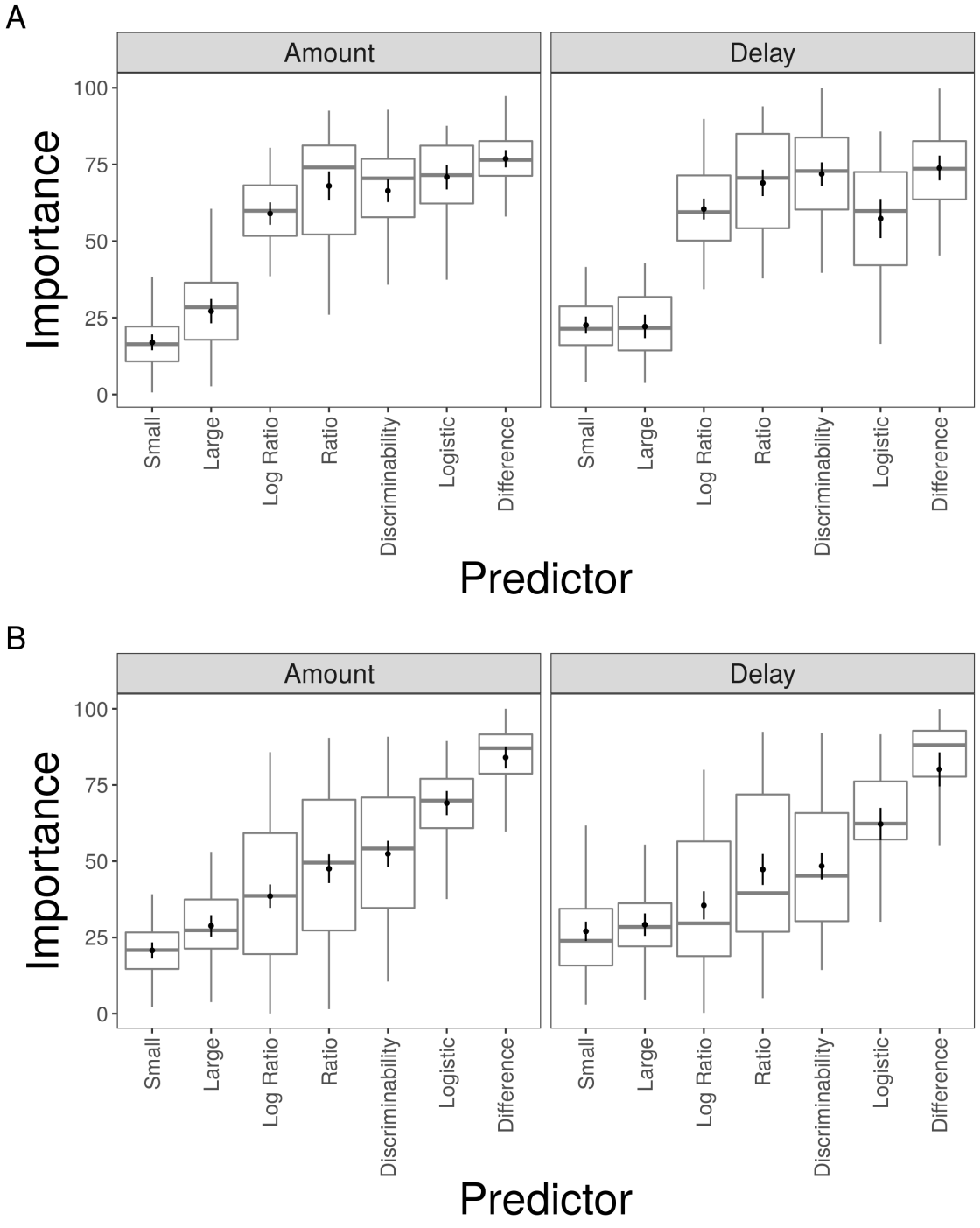

Figure S4. Predictor importance for each judgment type and data set (A = Data set 1, B = Data set 2). Predictor importance refers to the relative contribution of each predictor to the response. Predictors are ordered by overall mean importance. Dots represent means, error bars represent within-subjects 95% confidence intervals, boxplot horizontal lines represent medians, boxes represent interquartile range, whiskers represent 1.5 × interquartile range. Outliers are not shown. Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.

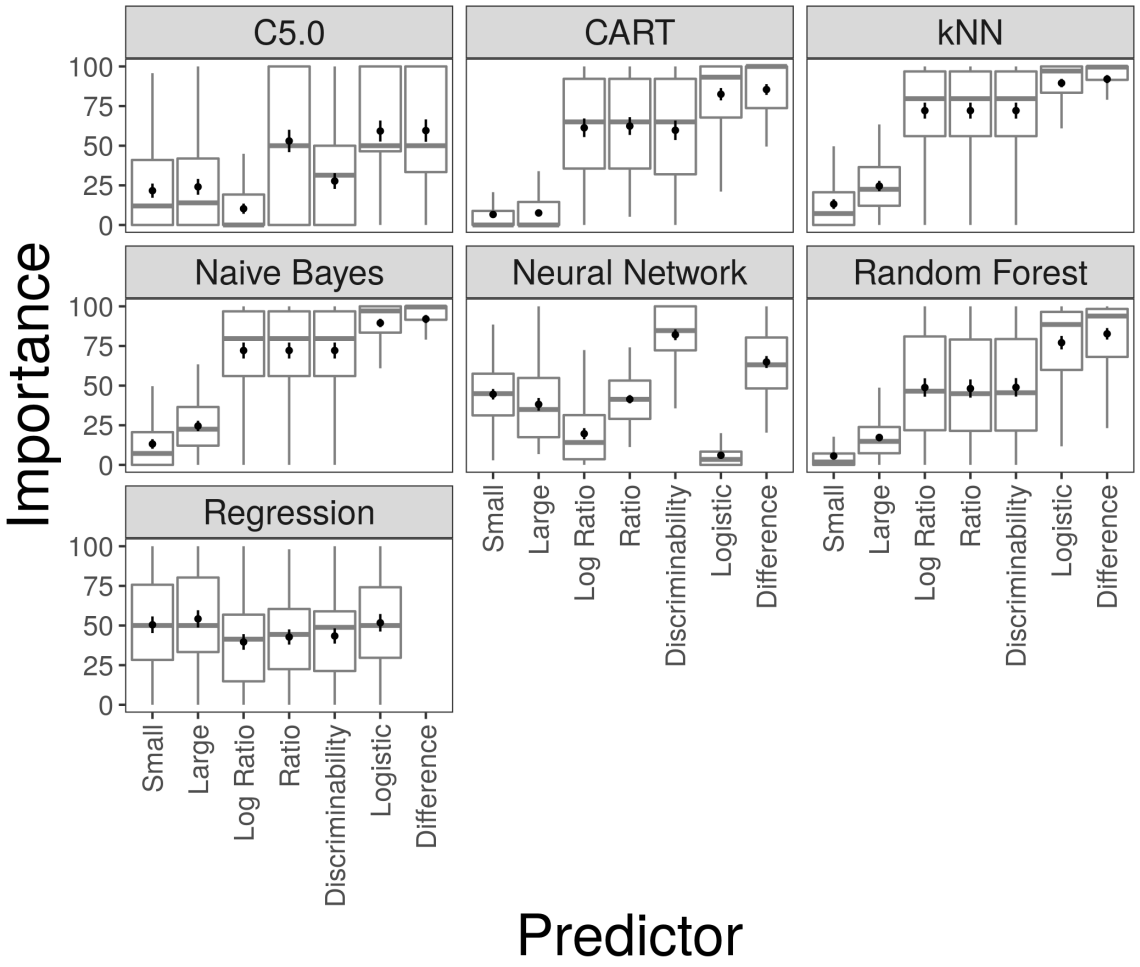

Figure S5. Predictor importance for each judgment type and algorithm. Predictor importance refers to the relative contribution of each predictor to the response. Predictors are ordered by overall mean importance. Dots represent means, error bars represent between-subjects 95% confidence intervals (the failure of regression models to calculate importance for the difference predictor prevents calculation of within-subject confidence intervals), boxplot horizontal lines represent medians, boxes represent interquartile range, whiskers represent 1.5 × interquartile range. Outliers are not shown. Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.

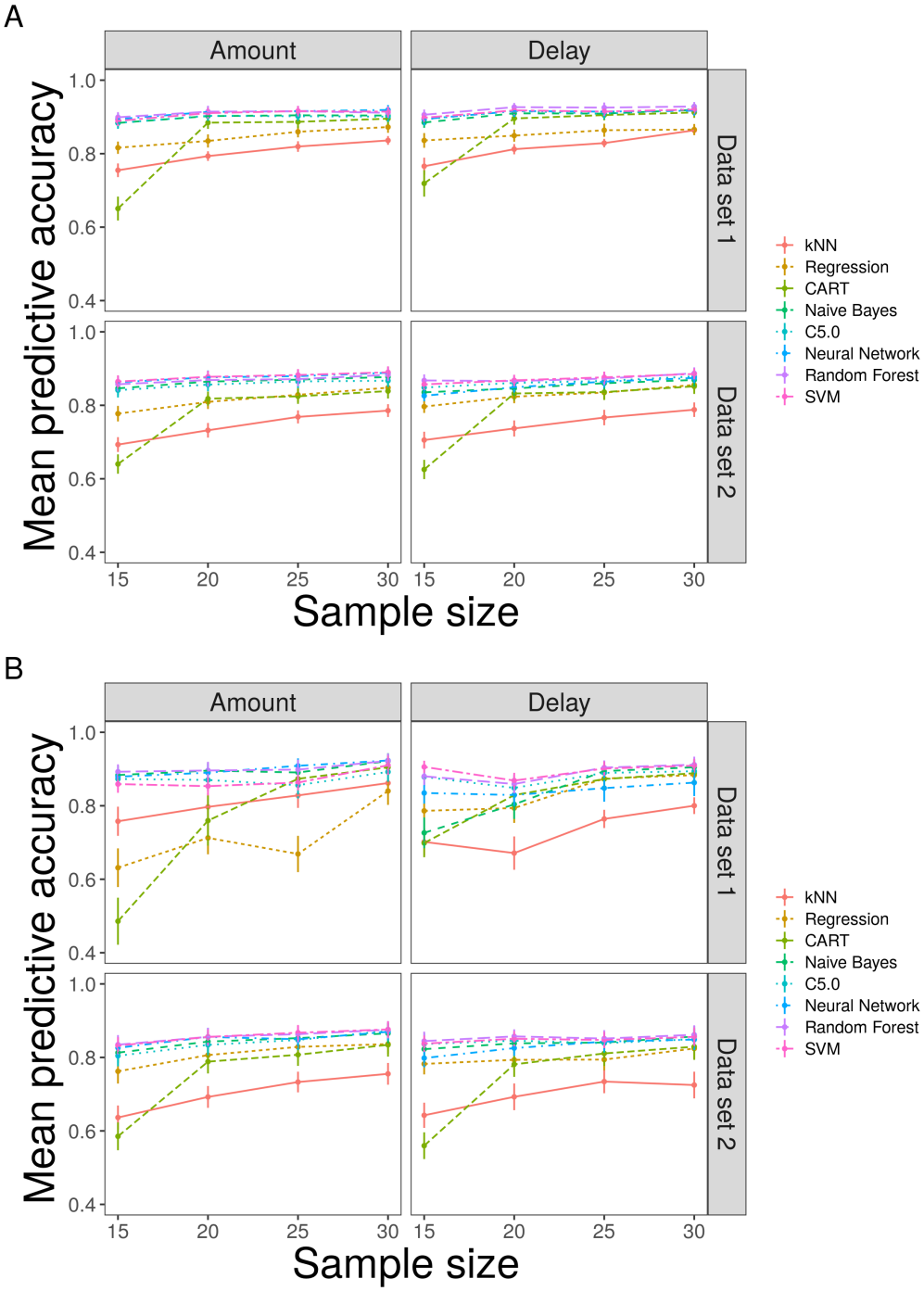

Figure S6. Mean predictive accuracy for each sample size, judgment type, data set, and ordering (A = Random, B = Sequential). Sample size refers to number of questions per participant used to train the algorithms. Random refers to a random sample of training questions used to predict a random sample of 10 testing questions. Sequential refers to a sample of training questions drawn in order of presentation to each participant that was used to predict a random sample of 10 testing questions. Dots represent means, and error bars represent between-subjects 95% confidence intervals (within-subject confidence intervals were not used because excessive missing data for small sample sizes caused too many participants to be removed from the calculations). Figure used with permission under a CC-BY4.0 license: Stevens et al., 2020; available at https://doi.org/10.17605/OSF.IO/WYTD9.