2018

Predicting Similarity Judgments in Intertemporal Choice with Machine Learning

Jeffrey R. Stevens
University of Nebraska-Lincoln, jstevens5@unl.edu

Leen-Kiat Soh
University of Nebraska, lsoh2@unl.edu

Predicting similarity judgments in intertemporal choice with machine learning

Jeffrey R. Stevens[1] & Leen-Kiat Soh[1]

[1] University of Nebraska-Lincoln

Abstract

Similarity models of intertemporal choice are heuristics that choose based on similarity judgments of the reward amounts and time delays. Yet, we do not know how these judgments are made. Here, we use machine-learning algorithms to assess what factors predict similarity judgments and whether decision trees capture the judgment outcomes and process. We find that combining small and large values into numerical differences and ratios and arranging them in tree-like structures can predict both similarity judgments and response times. Our results suggest that we can use machine learning to not only model decision outcomes but also model how decisions are made. Revealing how people make these important judgments may be useful in developing interventions to help them make better decisions.

Keywords: classification tree, decision tree, intertemporal choice, judgment, machine learning, similarity

Word count: 4118

Introduction

Would you prefer to receive \$100 today or \$105 in one month? Intertemporal choices such as these involve trading off smaller rewards available sooner with larger rewards available later. The temporal discounting approach to intertemporal choice models these tradeoffs by assuming that people subjectively devalue future rewards based on the time delay to receiving those rewards. Therefore, discounting models integrate the reward amount with the time delay to generate a discounted value for each option.

Jeffrey R. Stevens, Department of Psychology, Center for Brain, Biology & Behavior, University of Nebraska-Lincoln; Leen-Kiat Soh, Department of Computer Science and Engineering, University of Nebraska-Lincoln.

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

Though discounting models have dominated intertemporal choice modeling efforts for decades, recent work has offered alternative, heuristic models (Scholten & Read, 2010; Ericson, White, Laibson, & Cohen, 2015). One alternative uses similarity judgments to make intertemporal choices (Leland, 2002; Rubinstein, 2003; Stevens, 2016). This model generates judgments of similarity for both the reward amounts (e.g., Is \$100 similar to \$105?) and the time delays (e.g., Is receiving something now similar to receiving it in one month?). If one of these is judged as similar but the other as dissimilar, then people choose based only on the dissimilar one. This can be modeled as a decision tree that inputs the similarity judgments and outputs a choice (Figure 1a). In the example above, the amounts may be judged as similar, whereas the delays are judged as dissimilar, so people choose based on the delays and opt for the smaller, sooner option. This sequential comparison of similarity judgments recruits a completely different set of cognitive processes than the value integration of discounting approaches.

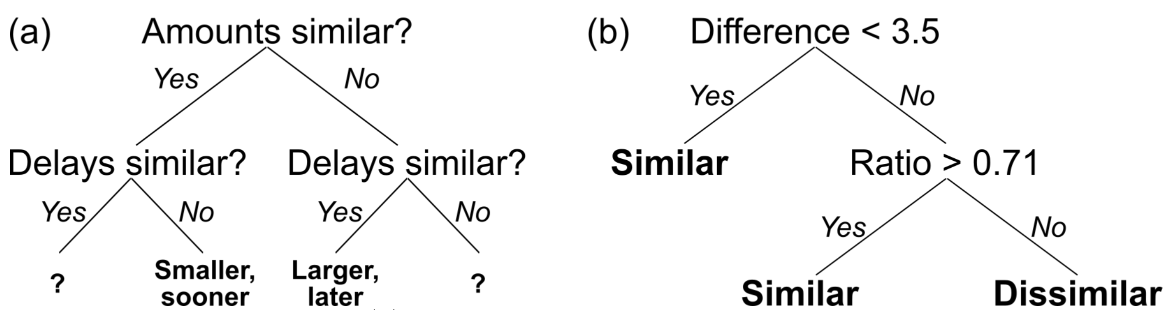

Figure 1. Similarity trees. (a) A similarity-based decision tree uses similarity judgments to make intertemporal choices. If amount or delay is judged as similar and the other as dissimilar, a choice is predicted. If both are similar or dissimilar, another choice rule must be used. (b) A decision tree can also be built to predict similarity judgments from combinations of small and large amount or delay values. This example illustrates that a judgment can be made at the first node (if the difference between values is < 3.5, judge as similar) or after a second node.

Behavioral data support the use of similarity judgments in intertemporal choices (Rubinstein, 2003; Stevens, 2016). In particular, Stevens (2016) measured similarity judgments and intertemporal choices and found that models incorporating these similarity judgments better predicted intertemporal choices than discounting models. But we do not know how these judgments are made: What makes \$3 vs. \$4 similar but \$3 vs. \$7 dissimilar? The aim of this study is to determine how people make similarity judgments and answer two key research questions:

1. *How do the small and large values of the reward amounts and time delays combine to predict similarity judgments?* Rubinstein (1988) proposed that either the numerical difference (large value − small value) or numerical ratio (small value / large value) between values could be used to make similarity judgments. For example, when comparing \$3 vs. \$4, one could focus on the difference of 1 or the ratio of 3/4. Stevens (2016) measured similarity judgments and found that both difference and ratio independently accounted for these judgments. Here, we test whether different mathematical operations combine small and large values to predict similarity judgments. We use classification algorithms from machine

learning to predict people's similarity judgments based on numerical difference and ratio or other psychophysical and decision-making functions (Table 1). This will tell us how small and large values combine to generate similarity judgments.

2. *Do trees capture similarity judgments?* Researchers often use regression models to investigate what factors classify responses. We propose an alternative classification method used in machine learning: *classification trees* (Breiman, Friedman, Olshen, & Stone, 1984). These algorithms produce *decision trees*, which are sequential decision rules for classifying outcomes based on a set of predictors. These trees are represented by nodes for each relevant predictor (e.g., difference or ratio) and a threshold for each predictor that divides into branches (Figure 1b). One can move down a tree by determining if the threshold of a predictor for a particular pair of values (e.g., $3 vs. $4) is met. Eventually, the tree ends in a terminal node that classifies the response. An advantage of decision trees is that they can make predictions not only for outcome data (e.g., choices, judgments) but also for process data (e.g., response times), which is useful for assessing decision strategies. In this study, we evaluate whether decision trees produced by machine-learning algorithms can model how similarity judgments are made by predicting both the judgment outcomes and response times.

To explore these questions, we used classification-tree algorithms from machine learning to assess what predictors best accounted for participants' similarity judgments and whether the resulting decision trees predicted judgments better than regression analyses. Combined, these findings reveal what cognitive processes influence similarity judgments.

Table 1
Predictors.

Predictor Name	Value/Function	Source
Small value	S	
Large value	L	
Difference	$L-S$	Rubinstein (1988)
Ratio	$\frac{S}{L}$	Rubinstein (1988)
Mean ratio	$\frac{S}{\frac{S+L}{2}}$	Eisler & Ekman (1959)
Log ratio	$\log(\frac{S}{L})$	Künnapas & Künnapas (1974)
Relative difference	$\frac{L-S}{L}$	González-Vallejo, Reid, & Schiltz (2003)
Disparity ratio	$\frac{L-S}{\frac{S+L}{2}}$	Boysen, Berntson, Hannan, & Cacioppo (1996)
Salience	$\frac{L-S}{S+L}$	Bordalo, Gennaioli, & Shleifer (2012)
Discriminability	$\log(\frac{L}{L-S})$	Welford (1960)
Logistic	$\frac{1}{1+e^{L-S}}$	

Methods

Data Sets

We tested our research questions on two data sets. Data set 1 was collected from 65 participants (29 males and 36 females) with a mean±SD age of 30.3±9.1 (range 22-72) years recruited from the Adaptive Behavior and Cognition Web Panel at the Max Planck Institute for Human Development in Berlin, Germany in August 2011. Participants received a flat fee of €3 for completing the survey. Web panel participants made similarity judgments between 50 pairs of amount values (e.g., €6 vs. €8) and 50 pairs of delay values (e.g., 6 days vs. 8 days): "Please decide whether the numbers are similar". This research was approved by the Max Planck Institute for Human Development's Ethics Committee.

Data set 2 was collected from 90 participants (29 males and 61 females) with a mean±SD age of 20.0±1.6 (range 18-26) years recruited from the University of Nebraska-Lincoln Department of Psychology undergraduate participant pool in December 2014. Participants received course credit for their participation. Participants started by making 20 intertemporal choices before rating the similarity of 43 reward amount values 43 and time delay values: "Do you consider receiving [small amount] and [large amount] to be similar or dissimilar?" and "Do you consider waiting [short delay] and [long delay] to be similar or dissimilar?". The intertemporal choices used the same value pairs as the similarity judgments and were included first to expose participants to the range of amount and delay magnitudes and to provide the overall decision context before they made similarity judgments. This research was approved by the University of Nebraska-Lincoln Internal Review Board (IRB Approval # 20130313118EP).

We chose the sample sizes of 65 and 90 because they were comparable to or greater than the sizes used in Stevens (2016), which detected medium-sized effects in the intertemporal choice model selection analyses. For both data sets, we recorded the similarity judgments for each question and demographic information, including age and gender. For data set 2, we also recorded response time and included attention checks with the same small and large value (10 vs. 10) or with very large differences between large and small values (1 vs. 90).

Classification Trees

Prior to the classification-tree analysis, we removed participants who (1) made the same similarity judgment in over 95% of the trials, (2) judged 10 vs. 10 to be dissimilar, (3) judged 1 vs. 90 to be similar, or (4) showed inconsistencies in judgments. To measure for inconsistencies, we included sets of questions in which the large value was fixed and was paired with at least 10 different small values. We removed participants with more than three switches between dissimilar to similar in at least one of these sets. In all, we removed 31 of the 155 participants, leaving 124 (Data set 1: $n = 50$; Data set 2: $n = 74$).

We used the machine-learning algorithm CART (Classification And Regression Trees; Breiman, Friedman, Olshen, & Stone, 1984) to classify similarity judgments. CART sequentially divides up data into groups based on predictor values to most accurately classify the data according to the response variable (for overview, see Loh, 2011). 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. This produces overly large trees that can overfit the data because the final groups must not have any 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. We limited the number of levels of nodes to three. Figure 2 illustrates trees and data from three example participants with different trees produced by CART.

We included a set of 11 predictors of similarity judgments (Table 1; Figure S1) for both CART and multiple logistic regression models. To compare the model classes, we used cross-validation to calculate out-of-sample predicted accuracy—the proportion of out-of-sample judgments accurately classified by the models. First, we randomly split the data in half (training sample and test sample). We then fit each model with all predictors on the training sample, which generated model-specific parameters (regression weights for each predictor and decision nodes and thresholds). Next, we used the fitted parameters to classify the test sample, which allowed us to calculate out-of-sample predicted accuracy. Finally, we switched the training and test samples and repeated the process. Model prediction occurred for each of the participants' data individually and separately for amounts and delays. Each participant's data was cross-validated 100 times for both decision-tree and regression models.

Data Analysis

For response time data, we removed outliers with modified Z scores greater than 3. We calculated Bayes factors (BF) to provide the weight of evidence for the alternative hypothesis relative to the null hypothesis (Wagenmakers, 2007). For example, $BF = 10$ means that the evidence for the alternative hypothesis is 10 times stronger than the evidence for the null hypothesis. Bayes factors between 1-3 provide only anecdotal evidence, those between 3-10 provides moderate evidence, those between 10-100 provide strong evidence,

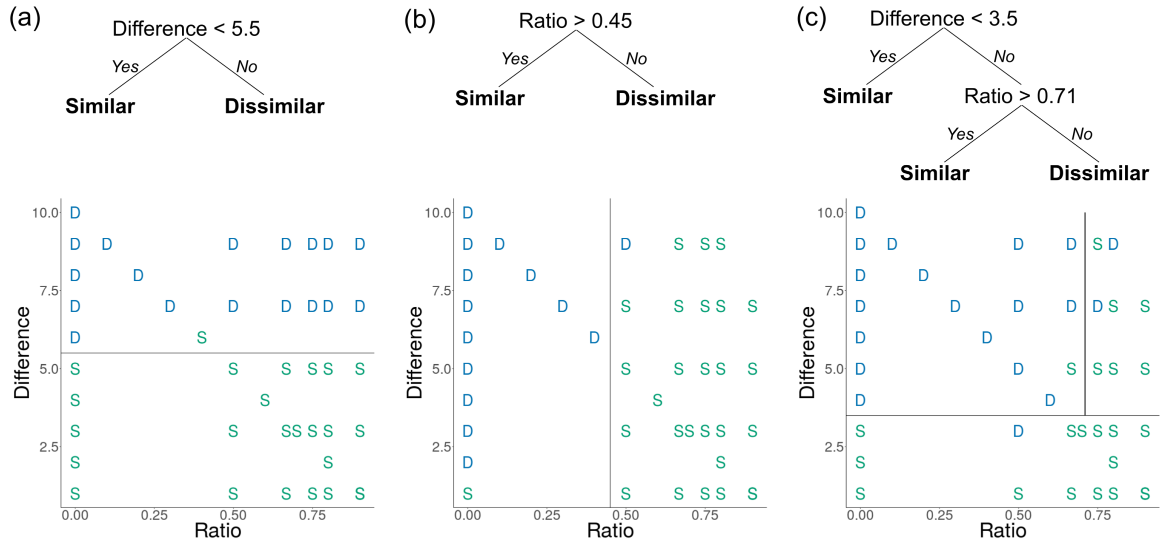

Figure 2. Decision trees and delay similarity judgments as a function of difference and ratio for example participants. Plots show individual value pairs coded by judgment (S=similar, D=dissimilar) as a function of difference and ratio of the value pairs. Horizontal lines represent difference thresholds. Vertical lines represent ratio thresholds. (a) This participant used only difference as a predictor, with a threshold of 5.5. This tree clearly classifies judgments quite well, with only one classification error (one similarity judgment for a value pair with a difference greater than 5.5) (b) This participant used only ratio as a predictor, with a threshold of 0.45 and two classification errors. (c) This participant used difference (threshold of 3.5) then ratio (threshold of 0.71) as predictors, with four classification errors.

and those above 100 provide very strong evidence (Andraszewicz, Scheibehenne, Rieskamp, Grasman, Verhagen, & Wagenmakers, 2015). Bayes factors associated with generalized linear mixed models were converted from Bayesian Information Criterion (BIC) using $BF = e^{\frac{BIC_{null} - BIC_{alternative}}{2}}$ (Wagenmakers, 2007). Bayes factors for t-tests were computed using noninformative priors (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

When comparing measures within a participant, we calculated within-subjects 95% confidence intervals (Morey, 2008). For mixed-effects models, we calculated profile likelihood 95% confidence intervals for coefficients. Confidence intervals are presented in brackets after the parameter estimate.

We analyzed the data using R Statistical Software version 3.4.2 (R Core Team, 2017)[1]. Data, R code, and supplementary tables and figures are available in the Supplementary Materials and at the Open Science Framework (https://osf.io/ew8dc/).

[1]We also used the `BayesFactor`, `car`, `cowplot`, `dplyr`, `foreach`, `ggplot2`, `lattice`, `lme4`, `MBESS`, `papaja`, `plyr`, `rpart`, `rpart.plot`, `tidyr`, and `xtable` packages (package usages and citations are provided in Supplementary Materials).

Results

Predictors of Similarity Judgments

Stevens (2016) demonstrated that both difference and ratio independently influence similarity judgments. Here, we (1) attempt to replicate this finding on new data and (2) evaluate how difference and ratio combine to predict similarity judgments. To address this, we restricted our analysis to data set 2, where we specifically created value pairs that varied difference while holding ratio constant and vice versa.

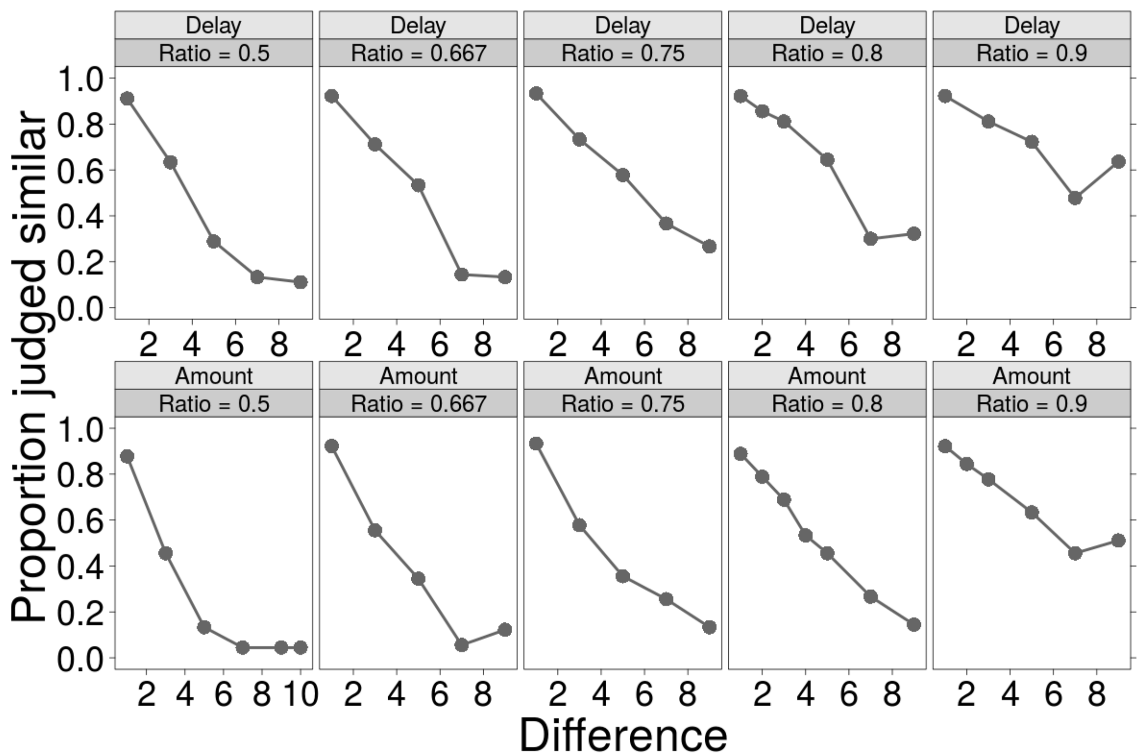

Figure 3. Difference and ratio effects on similarity judgments of amounts and delays in data set 2. Each panel represents the mean proportion of trials that participants judged value pairs to be similar for a given numerical ratio (0.5, 0.667, 0.75, 0.8, 0.9) and judgment type (amount or delay). The x-axis is the numerical difference between the value pair. Similarity judgments depended on both difference and ratio.

Figure 3 illustrates that difference and ratio both independently influence similarity judgments, replicating Stevens (2016). To explicitly test this, we conducted a binomially distributed generalized linear mixed model (GLMM) with similarity judgments as binary responses (0 for dissimilar, 1 for similar). We included difference, ratio, and judgment type (amount or delay) as fixed effects and participants as a random effect. Though we included the ratio × difference interaction, we did not include interactions between type and ratio or difference because we did not have a priori reasons to expect interactions and we wanted to test the simplest model possible. The GLMM confirmed that difference (β = -1.01 [-1.10, -0.91], $BF > 100$), ratio (β = 1.10 [0.51, 1.69], $BF > 100$), and type (β = 0.82 [0.68, 0.97],

Table 2
Best predictors for individual participant decision trees.

Data Set	Judgment Type	Large	Difference	Ratio	Relative Difference	Logistic
1	Amount	0	26	24	NA	NA
1	Delay	NA	25	25	NA	NA
2	Amount	0	62	10	1	1
2	Delay	1	50	15	3	2
All	Amount	0.0%	71.0%	27.4%	0.8%	0.8%
All	Delay	0.8%	62.0%	33.1%	2.5%	1.7%

Note. No participants had small, mean ratio, log ratio, disparity ratio, salience, or discriminability as the best predictor.

$BF > 100$) independently influenced similarity. Value pairs were judged as more similar with larger differences, with smaller ratios, and for delays compared to amounts. Furthermore, difference and ratio interacted ($\beta = 0.53$ [0.40, 0.66], $BF > 100$), with a weaker effect of difference at higher ratios. That is, as the ratio increased and values were more similar, the difference between values affected judgments less. People's judgments of similarity between two reward amounts or two time delays depended on *both* the numerical difference and numerical ratio. *Thus, both difference and ratio contributed to similarity judgments.*

The fact that both difference and ratio predict similarity judgments raises two possible causes. First, difference and ratio may combine mathematically, meaning they both are *simultaneously present* in the function used by our predictors (e.g., the predictor relative difference includes both difference and ratio in its expression—Table 1). Alternatively, difference and ratio may enter the tree *separately in sequence* (i.e., one predictor before the other one). We tested these alternative hypotheses by classifying similarity judgments with classification trees that included our predictors. If ratio and difference combine mathematically, then one of the combined predictors should best predict judgments for both amounts and delays. If they combine sequentially, then just difference and ratio predictors should be the best predictors of judgments.

For each participant and judgment type, the classification-tree algorithm generated a decision tree with the single best predictor for classifying the judgments (i.e., the first node in the tree). For 95-98% of participants across both data sets, either difference or ratio was the best predictor for amount and delay judgments (Table 2). *Thus, difference and ratio combined sequentially in a tree-like way to influence similarity judgments rather than in a more complicated mathematical operation.*

Table 3
Mean percent predicted accuracy for models

Judgment Type	Model	Mean Accuracy
Amount	Regression	80.1 [79.2, 81.1]
Amount	Tree	86.0 [84.9, 87.0]
Delay	Regression	77.4 [76.4, 78.5]
Delay	Tree	87.0 [86.1, 88.0]

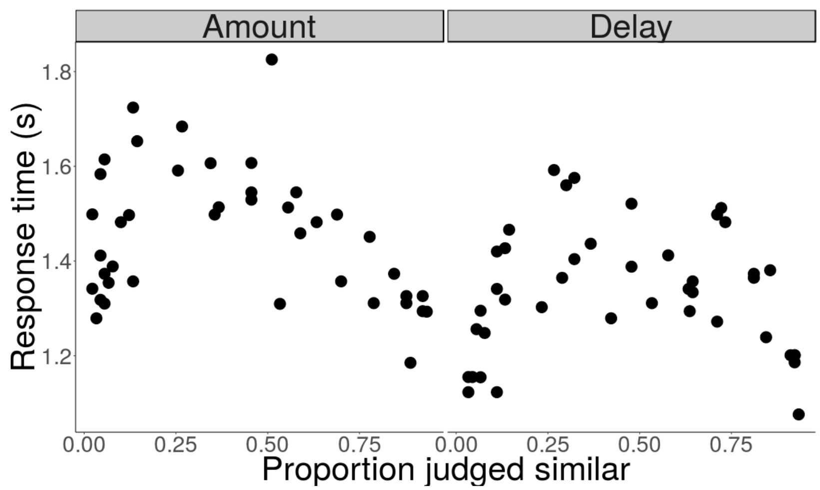

Figure 4. Similarity judgment effects on response time in data set 2. Each data point represents a value pair. The y-axis is the median response time for that pair. The x-axis is the mean proportion of participants judging that pair as similar.

Decision Trees as Process Models

Decision Trees Predict Similarity Judgments. To determine whether decision trees capture the outcome of making similarity judgments, we compared how both decision trees and regression models predicted similarity judgments for each participant's amount and delay judgments for both data sets. Decision trees outperformed regression models for out-of-sample predicted accuracy in amount judgments (Mean difference in accuracy = 5.8% [5.1, 6.6], Cohen's $d = 0.80$, $BF > 100$) and delay judgments (Mean difference in accuracy = 9.6% [8.5, 10.7], Cohen's $d = 1.09$, $BF > 100$) (Table 3). *Thus, decision trees predicted similarity judgments better than regression models.*

Decision Trees Track Response Time. Decision trees make predictions not only for judgment outcomes but also for aspects of the judgment process, namely response time, which we measured only in data set 2. Value pairs that are obviously similar or dissimilar should result in quick judgments. Intermediate value pairs, however, should be more difficult to judge, requiring longer response times. As expected, similarity judgments showed an inverted U-shaped relationship with response time for both amounts and delays (Figure 4), suggesting that value pairs with intermediate similarity judgments took more time to judge.

Decision trees may be able to track these differences in response time when judgments

Table 4
Number of participants trees with each number of nodes in data set 2

Judgment Type	1 Node	2 Nodes	3 Nodes	4 Nodes	5 Nodes
Amount	20	29	18	7	0
Delay	17	27	20	6	1

can be made after a single node or after multiple nodes (Figure 1b). If the judgment process follows a tree-like structure, we hypothesized that, when the tree predicts that the judgment requires traveling further down the tree, the participants' judgment times should increase due to processing multiple nodes. This was demonstrated in the fast-and-frugal priority heuristic for risky choices, where gambles that should only take one step to resolve had shorter responses times than gambles that took more than one step (Brandstätter, Gigerenzer, & Hertwig, 2006).

Participants varied in the number of nodes in their trees (Table 4). Those with two or more nodes allow for the possibility of stopping at different depths into the tree (node levels). Stopping at earlier node levels should result in shorter response times. Therefore, we restricted the analysis to participants in data set 2 whose trees allowed for stopping at different node levels as determined by CART (Amount: $n = 51$; Delay: $n = 52$; Figures S2 and S3). For each value pair, we determined at which decision node that participant's tree predicted that the judgment would be made. We then calculated the median response time for each participants' judgment at each node level and for each judgment type. We conducted a linear mixed effect model of median response time with number of node levels and judgment type as fixed factors and subject as a random factor (Figure 5). Number of node levels positively predicted response times ($\beta = 0.14$ [0.09, 0.20], $BF > 100$) but judgment type did not ($\beta =$ -0.14 [-0.30, 0.01], $BF = 0.24$), and there was no interaction ($\beta =$ 0.02 [-0.06, 0.10], $BF = 0.01$). Judgment response time, therefore, increased as participants had to work their way down the trees. *Thus, response time data were consistent with decision tree processing predictions.*

Discussion

Our results reveal that numerical difference and ratio predict similarity judgments for amounts and delays. Classification-tree algorithms indicate that, rather than combining mathematically, difference and ratio predictors are used separately and sequentially to make these judgments. These trees outperform regression models in predicting similarity judgments, and response time data suggest that decision trees not only predict judgment outcomes but also hint at tree-like judgment processes: People may evaluate one predictor before moving to a second if the first fails to result in a judgment.

For most participants, small and large values combine in rather simple ways via numerical differences and ratios to generate similarity (Table 2). Although both difference and ratio influence similarity judgments (Figure 3), they do so separately rather than via more complicated mathematical relationships. Thus, rather than previously proposed decision-making and psychophysical functions (Table 1), simple differences and ratios best

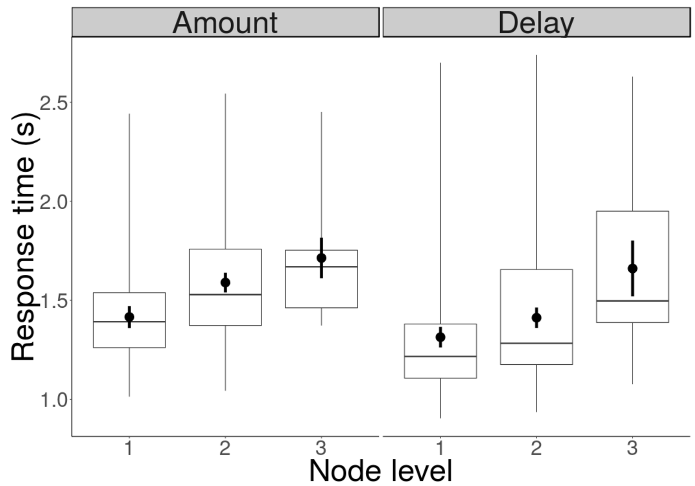

Figure 5. Response times as a function of decision tree nodes. Boxplots of participants' median response times show higher response times when decision trees predict the use of more node levels for both amount and delay judgments. Node level 3 includes judgments using three or more node levels (since there are so few participants with four or five node levels). Horizontal bars represent medians, boxes represent interquartile ranges, whiskers represent full ranges, dots represent means, and error bars represent within-subjects confidence intervals.

predict similarity judgments.

The importance of difference and ratio in similarity judgments mirrors patterns observed in psychophysical domains, including brightness, loudness, weight, and length (Stevens, 1975). Likewise, both difference and ratio are critical to human (and nonhuman) number discrimination. This is evidenced by the numerical distance effect, which shows discrimination based on difference (Rilling & McDiarmid, 1965), and Weber's law, which shows discrimination based on ratio (Mechner, 1958). Therefore, similarity judgments of monetary amounts and time delays follow core psychophysical principles of quantity judgments.

In this study, we used amount and delay magnitudes ranging from 0-100. Given that similarity judgments are context specific, the absolute magnitude of amounts and delays might influence how these judgments are made. First, the range of magnitudes assessed early on in testing might set anchors that bias judgments. We included the intertemporal choice questions before asking participants to make similarity judgments to illustrate the range of magnitudes and reduce bias and order effects. Second, participants may use different predictors, thresholds, or even classification algorithms across different magnitude ranges. Further work is needed to determine whether these results generalize across different magnitude ranges.

We also observed small differences in similiarity judgments across amount and delay judgment types (Figure 3; Table 2). While it is possible that these are meaningful differences, we do not yet have strong evidence that delay pairs are generally judged as more similar than amount pairs or that difference and ratio are better predictors for one judgment type over another. Further work is needed to investigate whehter there are robust differences between amount and delay judgments.

Rather than using classification-tree algorithms as only a statistical approach, we propose that these algorithms produce decision trees that might offer a process model of similarity judgments. Compared to regression models, decision trees use fewer predictors and compare predictor values to a threshold rather than weight them by a coefficient. Despite being simpler and more frugal in their information use, decision trees outperform regression models in predicting judgments.

Process data also support tree models: When decision trees predict the use of fewer nodes, participants indeed make judgments faster than when they are predicted to use more nodes. Both outcome and process data support decision trees as process models of similarity judgment. Since similarity judgments also apply to risky and strategic choice (Rubinstein, 1988; Leland, 2013), this approach can be extended to these choice domains, as well.

Understanding what factors influence similarity judgments is important because it provides opportunities to alter the "downstream" intertemporal choices. Therefore, these results not only give us insights into how people make these choices, but may also inspire interventions to help them make better decisions. Interventions that increase similarity judgments of time delays may focus attention on the reward amounts and nudge people into making more patient choices for their long-term benefit. This could help people improve their long-term health (diet, exercise, alcohol and drug consumption), financial stability (credit card debt reduction, retirement savings), and environmental sustainability (resource consumption, pollution reduction).

In conclusion, the similarity model can account for both outcome and process data in intertemporal choices (Leland, 2002; Rubinstein, 2003; Stevens, 2016), risky choices (Rubinstein, 1988; Leland, 1998), and strategic choice (Leland, 2013). This model moves the bulk of the decision process from the choice to the similarity judgment. Our work addresses how people make similarity judgments by showing that (1) rather simple combinations of small and large values (numerical differences and ratios) can predict similarity judgments and (2) decision trees capture both the outcome and process data. We used machine learning algorithms to not only create statistical models of judgment outcomes but also develop process models that capture how decisions are made. Thus, machine-learning algorithms provide a useful set of tools for modeling judgment and decision making, with the potential to help people make better decisions.

Acknowledgments

This research was funded by an Alexander von Humboldt Foundation TransCoop Grant and by National Science Foundation grants (NSF-1062045, NSF-1658837). We would like to thank Isabella Otto for collecting data in Germany; Duy Nguyen for developing the Java-based data collection program and helping analyze data; Nik Leger and Cherylynn Gibson for helping analyze data; Noah Svec for testing participants; and UNL's CB3 Club for comments on an early draft.

References

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521–543. doi:10.1177/0149206314560412.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, *127*(3), 1243–1285. doi:10.1093/qje/qjs018.

Boysen, S. T., Berntson, G. G., Hannan, M. B., & Cacioppo, J. T. (1996). Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes). Journal of Experimental Psychology: Animal Behavior Processes*, *22*(1), 76–86. doi:10.1037/0097-7403.22.1.76.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409–432. doi:10.1037/0033-295X.113.2.409.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman; Hall.

Eisler, H., & Ekman, G. (1959). A mechanism of subjective similarity. *Nordisk Psykologi*, *11*(1), 1–10. doi:10.1080/00291463.1959.10780403.

Ericson, K. M. M., White, J. M., Laibson, D. I., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. *Psychological Science*, *26*(6), 826–833. doi:10.1177/0956797615572232.

González-Vallejo, C., Reid, A. A., & Schiltz, J. (2003). Context effects: The proportional difference model and the reflection of preference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 942–953. doi:10.1037/0278-7393.29.5.942.

Künnapas, T., & Künnapas, U. (1974). On the mechanism of subjective similarity for unidimensional continua. *The American Journal of Psychology*, *87*(1/2), 215–222. doi:10.2307/1422015.

Leland, J. W. (1998). Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory. *Management Science*, *44*(5), 659–672. doi:10.1287/mnsc.44.5.659.

Leland, J. W. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry*, *40*(4), 574–581. doi:10.1093/ei/40.4.574.

Leland, J. W. (2013). Equilibrium selection, similarity judgments, and the "nothing to gain/nothing to lose" effect. *Journal of Behavioral Decision Making*, *26*(5), 418–428. doi:10.1002/bdm.1772.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. doi:10.1002/widm.8.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, *1*(2), 109–121. doi:10.1901/jeab.1958.1-109.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64. doi:10.20982/tqmp.04.2.p061.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed-ratio schedules. *Science*, *148*(3669), 526–527. doi:10.1126/science.148.3669.526.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi:10.3758/PBR.16.2.225.

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *46*(1), 145–153. doi:10.1016/0022-0531(88)90154-8.

Rubinstein, A. (2003). "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review*, *44*(4), 1207–1216. doi:10.1111/1468-2354.t01-1-00106.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review*, *117*(3), 925–944. doi:10.1037/a0019619.

Stevens, J. R. (2016). Intertemporal similarity: Discounting as a last resort. *Journal of Behavioral Decision Making*, *29*(1), 12–24. doi:10.1002/bdm.1870.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: John Wiley & Sons.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. doi:10.3758/BF03194105.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, *3*(3), 189–230. doi:10.1080/00140136008930484.