Psychological Forest: Predicting Human Behavior

Ori Plonsky,[1] Ido Erev,[2] Tamir Hazan,[3] Moshe Tennenholtz[4]

Technion - Israel Institute of Technology, Haifa, 3200003, Israel

[1]plonsky@campus.technion.ac.il [2]erev@tx.technion.ac.il [3]tamir.hazan@technion.ac.il [4]moshet@ie.technion.ac.il

Abstract

We introduce a synergetic approach incorporating psychological theories and data science in service of predicting human behavior. Our method harnesses psychological theories to extract rigorous features to a data science algorithm. We demonstrate that this approach can be extremely powerful in a fundamental human choice setting. In particular, a random forest algorithm that makes use of psychological features that we derive, dubbed psychological forest, leads to prediction that significantly outperforms best practices in a choice prediction competition. Our results also suggest that this integrative approach is vital for data science tools to perform reasonably well on the data. Finally, we discuss how social scientists can learn from using this approach and conclude that integrating social and data science practices is a highly fruitful path for future research of human behavior.

1 Introduction

Prominent speech recognition researcher Fred Jelinek is often quoted for saying "Every time I fire a linguist, the performance of our speech recognition system goes up". This saying highlights a common wisdom in data science according to which social scientists - and their theories - are of little help when it comes to the development of useful data analytic tools. In sharp contrast with this wisdom, choice prediction competitions (tournaments aimed for prediction of human behavior) organized by social scientists (Erev et al. 2010), show a large advantage of models that build on social science theories over data-based computational tools.

We believe this apparent inconsistency is a result of improper or insufficient integration between the disciplines. The main goal of the current paper is to demonstrate the merits of integrating data science and social science, in the context of predicting human choice behavior. In this domain, common practices - and the current state-of-the-art - either (a) focus purely on the data scientific tools, mostly neglecting insights from the choice psychology literature; or (b) focus on the psychological drivers of choice, amalgamating them only heuristically, rather than rigorously. Our study aims to bridge this gap by developing computational tools fed with features derived from psychological theories. That is, first, we use psychological theories to identify potentially

relevant features. Then, we allow the computational tools to decide on the best manner by which these features are to be integrated to a unified model (see a similar idea developed independently by Noti et al., 2016). Note this approach differs from work that integrates data science and psychological theories by assuming or using existing psychological models and then designing computational agents that respond to these models (Azaria et al. 2012a; 2012b; De Melo, Gratch, and Carnevale 2014; Prada and Paiva 2009). We compare the predictive accuracy of our approach with current practice and show it outperforms the state-of-the-art for our data.

Additionally, this integrative approach provides benefits for both social scientists and data scientists. Data scientists working along a wide array of domains related to choice behavior may often seek to harness psychological insights to improve the performance of their predictive tools. Yet, they face two main challenges. First, psychological theories abound and it is unclear which one is most suitable for a specific task. We take up on this challenge by studying a domain that is often considered to encapsulate the most basic tenets of human choice behavior. Therefore, many insights the study of this domain provides should generalize to other domains as well. A second challenge is that translation of the theory to specific meaningful elements that can be useful for the development of predictive tools is rarely straightforward. Here, we engineer clear features that can be seamlessly used across domains. Importantly, we demonstrate the efficacy of these features by showing that their use in data science algorithms significantly improves upon best performance achieved by learning algorithms trained without them.

Social scientists can also be informed from our integrative approach. Models social scientists develop require making auxiliary assumptions regarding the exact implementation of the interactions between the various underlying theoretical elements. A test of these models is then simultaneously a test of both the theory and the auxiliary assumptions made. In contrast, derivation of clear features from the theory allows a construction of multiple learners based on the theory using existing data science tools. These, in turn, can be easily tested and thus an examination of the theoretical building blocks can be disentangled from that of the auxiliary assumptions the modeler makes. Moreover, many algorithms also provide the benefits of discovering which of the underlying features

is most important, which then informs the social scientist which of the theoretical constructs is indeed most relevant.

2 Choice Prediction Tasks

We focus on the prediction of aggregate human choice behavior over time. Consider for example an insurance company that offers a discount for drivers willing to use an In-Vehicle Data Recorder. The size of the discount is reduced when reckless driving is recorded. The company considers two incentives schemes. The first deducts from the discount $0.1 for each recorded safety violation; the second deducts $10, but only with .01 probability for each violation. To support the choice between the two schemes, the company wishes to predict the frequency of violations given each scheme over time. Notice that though the company may have data regarding past behavior in different scenarios (e.g. drivers' response to traffic enforcement cameras) from which it can learn, to predict behavior in the current novel setting, it is likely that it should also leverage on some theory regrading the mechanisms impacting the drivers' decisions.

Clearly, many aspects can affect people's particular driving decision. To learn more general aspects of choice behavior that can be generalized across domains, we focus on a more abstract domain, choice between gambles. Choice between gambles has become a drosophila of human decision research and is one of the best studied topics in behavioral economics (Kahneman and Tversky 1979; Savage 1954; Tversky and Kahneman 1992). It is commonly used as a proxy for human preferences over economic products (Golovin, Krause, and Ray 2010; Srivastava, Vul, and Schrater 2014) and its study assumes it reveals basic human attitudes toward risk and value (Savage 1954; Kahneman and Tversky 1984). Moreover, many classical behavioral phenomena have been originally demonstrated using this paradigm (Allais 1953; Kahneman and Tversky 1979). Furthermore, in this domain social theories were found to be particularly useful in choice prediction competitions (CPC), challenges for the prediction of choice decisions made by humans in controlled lab experiments. Therefore, the focus on this domain allows both a solid theoretical framework for development of psychological features and strong benchmarks to compare our learners to.

Specifically, the data we use here was collected as part of a recent CPC (Erev, Ert, and Plonsky 2015)[1] aimed for development of predictive tools for choice between gambles over time. Thus, it is focused not only on the initial decisions people make when facing gambles, but also on the development of their choices over time after obtaining feedback regarding these gambles. Unlike most time-series decision research however, the CPC focuses on predictions of the mean time-dependent choice rates for the whole series in advance. That is, predictions made for time t cannot use the actual choices made before time t. This type of content filtering task can simulate, for instance, the attempt to predict the development over time of the public response to two possible products or to two possible policies, like the decision made by the insurance company above.

In each choice problem in this CPC, two gambles, A and B, are displayed to a decision maker, which is then asked to choose between them repeatedly, for 25 times. After each decision, a computer draws two outcomes, one for each gamble, in accordance with the gambles' payoff distributions in that choice problem. The payoff drawn for the chosen alternative is then the payoff the decision maker obtains for that decision. Yet, in the first five decisions of each problem, the decision maker does not get feedback regarding this (or any other) payoff. Following each of the other decisions (i.e. as of the 6th choice), the decision maker gets complete feedback regarding the drawn outcomes: Both the obtained payoff and the forgone payoff (the outcome of the non-chosen gamble) for that decision are presented. Note that the information initially provided for the two gambles remains on-screen for all 25 decisions, and only the feedback provided may change from one decision to another.

Each five consecutive decisions are called a *block*. Thus, each choice problem contains five blocks: the first of decisions made without feedback and the rest of decisions made with (and following) feedback. The goal of the participants in the CPC was to predict the mean aggregate (across all participants) choice rate of one of the gambles in each block and in each problem.

3 Features of Choice Problems

Each choice problem in our data is a repeated decision between a pair of gambles (A, B), and is uniquely defined by 11 parameters. The distribution of A, $F_A = (A_1, q_1; A_2, 1 - q_1)$ is defined by the three parameters A_1, q_1, A_2. The distribution of B, $F_B = (B_1, p_1; B_2, p_2, ...B_m, 1 - \sum_{i=1}^{m-1} p_i), m \leq 10$, is defined by the five parameters $(B_1, p_1, LotVal, LotShape, LotNum)$, where the latter three define a lottery (provided with probability $1 - p_1$) which sets the values of $\{(B_i, p_i)\}_{i=2}^m$. $LotVal$ is the lottery's expected value (EV), $LotShape$ is its distribution shape (symmetric, right-skewed or left-skewed), and $LotNum$ is its number of possible outcomes. A ninth parameter, Amb, defines whether Gamble B is ambiguous. If the gamble is ambiguous, then the probabilities of the possible outcomes, $\{p_i\}$, are undisclosed to the decision maker (they are replaced with the symbols $p_1, ..., p_m$). A tenth parameter, $Corr$, captures the correlation between the outcomes that the two gambles generate (either positive, negative or none). The 11th parameter, $Feedback$, captures whether feedback is provided to the decision maker. As explained above, it is set to 0 in the 1th block of each problem, and to 1 in all other blocks. Each of the 11 parameters that define a problem is provided explicitly to the decision makers in some way. For example, decision makers see a full description detailing the payoff distributions of both gambles (unless Gamble B is ambiguous) and are told whether a correlation between the two gambles exists.

To compare the usefulness of adding psychological insights to computational tools, we define three feature sets and use them for the development of the different learners

---

[1] The competition's website: http://departments.agri.huji.ac.il/cpc2015

Table 1: Feature sets used in the experiments

Set name	Features included
Objective	$A_1, q_1, A_2, B_1, p_1, LotVal, LotShape, LotNum, Amb, Corr, Feedback$
Naïve	$dEV, dSD, dMins, dMaxs$
Psychological	$dEV_0, dEV_{FB}, pBetter_0, pBetter_{FB}, dUniEV, pBetter_U, dSignEV, pBetter_{S0}, pBetter_{SFB}, dMins, SignMax, RatioMin, Dom$

In addition, a $block$ feature is added to each set.

to be tested. The first, *Objective* feature set is void of theory, and includes the 11 parameters that define each problem with an additional $block$ feature that captures the development of choice over time (i.e. equals $1, 2, ...5$; see Table 1). The second, *Naïve* feature set, includes four domain-relevant features that can serve as a reasonable starting point capturing domain knowledge. They capture very basic properties of the choice problem and represent basic decision rules according to which humans can make a decision. The four features are dEV, the difference between the gambles' objective expected values; dSD, the difference between the gambles' standard deviations; $dMins$, the difference between the gambles' minimal outcomes; and $dMaxs$, the difference between the gambles' maximal outcomes. We consider these domain-relevant features naïve because most modelers of human choice data are likely to test these features even without knowledge of any psychological theory.

Psychological Features. The third, *Psychological* feature set, includes 13 features that aim to capture directly research made by social scientists on decision making and the psychology of choice. The first psychological features are motivated by the observation that in choice between gambles, decision makers tend to be sensitive to the difference between the gambles' expected values (EVs) (Erev and Haruvy 2016). However, in ambiguous choice problems (under which decision makers cannot compute the EV of Gamble B), the difference between the EVs needs to first be estimated by the decision makers. Specifically, previous behavioral research suggests that when facing ambiguity, decision makers: (a) tend to be pessimist with respect to their available outcomes (Gilboa and Schmeidler 1989; Wakker 2010), (b) tend to have a flat prior regarding the possible outcomes (Viscusi 1989), and (c) assume that the alternative option can serve as a reasonable approximation for the value of the ambiguous option (i.e., assume that the EVs are not likely to be very different) (Garner 1954). Moreover, previous research also suggests that feedback leads choice towards the actual EV (Erev and Barron 2005). Therefore, two features relating to the difference between the EVs, or the estimate of this difference, are introduced to the psychological set: one capturing an estimate of Gamble B's EV before feedback and another capturing this estimate with feedback:

$$dEV_0 = \begin{cases} EV_B - EV_A & \text{non-amb.} \\ (Min_B + UEV_B + EV_A)/3 - EV_A & \text{ambiguous} \end{cases} \quad (1)$$

$$dEV_{FB} = \frac{1}{2}(dEV + dEV_0) \quad (2)$$

where Min_B is the minimal possible outcome of B (thus providing more weight to the worst outcome and introducing pessimism) and UEV_B is the EV of B when all outcomes are equally likely.

Although fairly sensitive to the difference between EVs, vast behavioral research shows that other, somewhat less normative aspects of the choice problem, also influence the decisions people make (Kahneman and Tversky 1979; Brandstätter, Gigerenzer, and Hertwig 2006). First, it has been suggested that people try to minimize immediate regret by preferring the option that leads to a better outcome most of the time (Erev and Roth 2014). We introduce two features that capture the (estimated) probability that one gamble generates a better (higher) outcome than the other. Specifically, before obtaining feedback, decision makers may try to estimate this probability from the available description by (mentally) comparing the gambles' distributions; and after obtaining feedback, they may simply use the observed frequency of trials at which one gamble was better than the other. In the latter case, their estimate also depends on the correlation between the outcomes that the gambles generate[2]:

$$pBetter_0 = P[F_B^{-1}(x_1) > F_A^{-1}(x_1)] - P[F_B^{-1}(x_1) < F_A^{-1}(x_1)] \quad (3)$$

$$pBetter_{FB} = \begin{cases} P[F_B^{-1}(x_1) > F_A^{-1}(x_1)] \\ - P[F_B^{-1}(x_1) < F_A^{-1}(x_1)] & Corr > 0 \\ P[F_B^{-1}(x_1) > F_A^{-1}(1 - x_1)] \\ - P[F_B^{-1}(x_1) < F_A^{-1}(1 - x_1)] & Corr < 0 \\ P[F_B^{-1}(x_1) > F_A^{-1}(x_2)] \\ - P[F_B^{-1}(x_1) < F_A^{-1}(x_2)] & Corr = 0 \end{cases} \quad (4)$$

where F^{-1} is the inverse cumulative distribution function and $x_i \sim U[0, 1]$.

Previous behavioral research also suggests that instead of using a cumbersome process of computing the EVs of the gambles, some decision makers use simple heuristics to make their choices. One such heuristic is to treat the gambles as if their distribution is *uniform*, that is, to neglect the described probabilities and assume all possible outcomes are equally likely (Thorngate 1980). This assumption defines two different distributions, $F_{UA} = (A_1, 1/2; A_2, 1/2)$ and $F_{UB} = (B_1, 1/m; ...; B_m, 1/m)$. Such relaxation then allows for a much easier computation of the difference between the (new) EVs, as well as an easier identification of the gamble that is more likely to provide the better outcome. Following this logic, we add the following two features:

$$dUniEV = EV_{UB} - EV_{UA} \quad (5)$$

---

[2]If the problem is ambiguous, the probabilities should first be estimated. Past research suggests that they should be estimated such that the minimal outcome is more likely than the others (incorporating pessimism) while all other outcomes are equally likely. To maintain consistency, the probabilities are also estimated such that the difference between the EV that their estimates imply and the estimated EV implied by Equation 1 is minimal.

$$pBetter_U = P[F_{UB}^{-1}(x) > F_{UA}^{-1}(x)]$$
$$- P[F_{UB}^{-1}(x) < F_{UA}^{-1}(x)], \quad x \sim U[0,1] \quad (6)$$

Another heuristic commonly used in choice between gambles is a *sign* heuristic, according to which the magnitudes of the possible outcomes are neglected and only the total probabilities of winning or losing are considered relevant (Payne 2005). Again, this heuristic defines two distributions: $F_{SA} = (sign(A_1), q_1; sign(A_2), 1 - q_1)$ and $F_{SB} = (sign(B_1), p_1; ...; sign(B_m), 1 - \sum p_i)$, where $sign(\cdot)$ is the sign transformation. Like the uniform heuristic, the new biased distributions imply sensitivity both to the estimated EVs and to the likelihood that one gamble provides a better outcome than the other. Yet, unlike the uniform heuristic, the actual probabilities are used here and thus the estimations of these probabilities in case of ambiguity is likely to change prior and with feedback (cf. Eq. 3, 4). Thus, three features are introduced here to the psychological set:

$$dSignEV = EV_{SB} - EV_{SA} \quad (7)$$

$$pBetter_{S0} = P[F_{SB}^{-1}(x) > F_{SA}^{-1}(x)]$$
$$- P[F_{SB}^{-1}(x) < F_{SA}^{-1}(x)], x \sim U[0,1] \quad (8)$$

$$pBetter_{SFB} =$$
$$\begin{cases} P[F_{SB}^{-1}(x_1) > F_{SA}^{-1}(x_1)] \\ \quad - P[F_{SB}^{-1}(x_1) < F_{SA}^{-1}(x_1)] & Corr > 0 \\ P[F_{SB}^{-1}(x_1) > F_{SA}^{-1}(1 - x_1)] \\ \quad - P[F_{SB}^{-1}(x_1) < F_{SA}^{-1}(1 - x_1)] & Corr < 0 \\ P[F_{SB}^{-1}(x_1) > F_{SA}^{-1}(x_2)] \\ \quad - P[F_{SB}^{-1}(x_1) < F_{SA}^{-1}(x_2)] & Corr = 0 \end{cases} \quad (9)$$

where $x_i \sim U[0,1]$.[3]

Another feature treated here is a *minimax* heuristic (Edwards 1954; Brandstätter, Gigerenzer, and Hertwig 2006). This heuristic prescribes choice of the gamble with the better (higher) minimal outcome. A feature capturing this tendency is added to the psychological set as well:

$$dMins = Min_B - Min_A \quad (10)$$

Yet, it has also been suggested that decision makers use this pessimistic strategy less when they feel it is futile to use it. Specifically, it is avoided more when, regardless of choice, the decision maker has no possibility to gain anything. This type of behavior leads to the so-called reflection effect (Markowitz 1952; Kahneman and Tversky 1979), according to which people are risk averse in the gain domain and risk seeking in the loss domain. To capture this possibility, we introduce a feature signaling whether gains are even possible:

$$SignMax = sign(max\{A_1, A_2, B_1, ..., B_m\}) \quad (11)$$

Alternatively, this pessimistic tendency may feel futile when the difference between the minimal outcomes is negligible. Thus, a feature capturing the ratio between the minimal outcomes is added:

$$RatioMin =$$
$$\begin{cases} 1 & Min_A = Min_B \\ \frac{min\{|Min_A|, |Min_B|\}}{max\{|Min_A|, |Min_B|\}} & Min_A \neq Min_B \wedge \\ & sign(Min_A) = sign(Min_B) \\ 0 & otherwise \end{cases} \quad (12)$$

---
[3]In Eq. 7, 8, in case of ambiguity, the probabilities are first estimated as explained in Footnote 2.

Finally, when choice problems are trivial, decision makers often recognize it and choose without performing unnecessary computations. Specifically, if one gamble stochastically dominates the other, the choice problem is trivial. Therefore, the final feature we add to the psychological set identifies whether one gamble dominates the other:

$$Dom = \begin{cases} 1 & [P(B \geq x) \geq P(A \geq x) \, \forall x] \wedge \\ & [\exists x : P(B \geq x) > P(A \geq x)] \\ -1 & [P(B \geq x) \leq P(A \geq x) \, \forall x] \wedge \\ & [\exists x : P(B \geq x) < P(A \geq x)] \\ 0 & otherwise \end{cases} \quad (13)$$

4 Experiments

Our experiments focus on the aggregate human choice behavior in different choice problems, and on its progression over time. To that end, we use the CPC data (available, with more detailed accounts, at the CPC's website). The data includes decisions made by 446 incentivized human participants in 150 different choice problems, which are all points in the same 11-dimensional space. In the CPC, 90 problems served as training data and the other 60 served as the test data. Thirty of the training problems were carefully-selected from the space because they pose special interest to decision researchers. The other 120 problems were randomly selected according to a predefined algorithm. Each decision maker faced 30 problems (of either only the train set or only the test set) and made 25 decisions divided to five blocks in each of these problems.

We compare the value of different learning algorithms, using various combinations of the feature sets above, in the task participants of the CPC faced: Prediction of both the initial aggregate choice behavior and its progression over time in novel (previously unobserved) settings. Thus, we train each algorithm-features combination on the CPC's training data of 90 choice settings (each consisting five time-points, or blocks) and test its predictive value in the CPC's test data of 60 different choice settings. Specifically, the variable of interest is the mean aggregate choice rate for one of the two alternatives in each block and in each setting. Performance is thus measured according to MSE of 300 choice rates in the range $[0,1]$

Table 2 shows four relevant benchmarks for the predictive performance in the current task. Benchmark *Random* predicts 50% choice for each alternative. Benchmark *Average* predicts, for each game and each block, the mean choice rate observed in the 90 train problems for that block. Benchmark *BEAST* refers the CPC's baseline model (dubbed BEAST), a purely psychological model developed by social scientists. The mechanics and underlying theory of BEAST were the di-

Table 2: Benchmark models

Benchmark	Performance ($MSE * 100$)
Random	7.62
Average	7.76
BEAST	0.99
CPC Winner	0.88

rect inspiration for the 13 psychological features introduced above. In that sense, the *Psychological* feature set can be thought of as if it contains building blocks of this baseline model (though note BEAST itself does not define these psychological features). Benchmark *CPC Winner* refers to the winning model of the current CPC. This model is a minor refinement of BEAST with an additional heuristic not discussed here. In particular, all psychologically-inspired features are components of the winner as well. Importantly, the winner's improved performance over BEAST is not statistically significant.

The algorithms tested include random forest (using R package *randomForest*); neural nets (using R package *neuralnet*) with one hidden layer and either 3, 6, or 12 nodes and with two hidden layers and either 3 or 6 nodes in each layer; SVM (using R package *e1071*) with radial and polynomial kernels; and kNN (using R package *kknn*) with 1, 3, or 5 nearest neighbors.

We trained each algorithm-features combination with both the packages' default hyper-parameter values and with values tuned to fit the training data best (according to 10 rounds of 10-fold cross validation). The qualitative results of both methods were very similar and we thus present only the results of the default values method. Note that off-the-shelf algorithms that do not require too much fine tuning are much more likely to appeal to researchers of other fields, like psychologists.

Results. Table 3 exhibits the results of the various algorithm-feature combinations in predicting behavior in the test set. It suggests that for the current data, no simple algorithm can have reasonable predictive performance without using the psychological features. In particular, best predictive performance for an algorithm using only the *Objective* and/or the *Naïve* features (random forest using both sets of features) reflects MSE of 0.0142, which is 61% worse than the predictive performance of the CPC winner, a purely psychological baseline developed by social scientists. To test for the robustness of this result, we computed, using a bootstrap analysis with 1000 replicates, a confidence interval for the difference in prediction MSEs between each learner and the benchmarks. The results suggest that each of the algorithms not using the *Psychological* feature set predicts significantly worse than both BEAST and the CPC winner.

The results also show that a simple random forest algorithm using only the *Psychological* feature set already slightly outperforms the baseline BEAST which inspired each of the features in this set. Moreover, adding more features to this algorithm improves its predictive performance, which suggests not all relevant features were components of the baseline model. Specifically, random forest using all three feature sets combined provides better predictive accuracy than the best previously available model developed for this data (the CPC winner). This type of *Psychological Forest* model achieves MSE of 0.0087, a relative improvement of 39% over the best algorithm not fed with any psychological features.

Interestingly, given almost any possible set of features used, random forests outperform all other algorithms. It is possible this is because random decision trees, with their

stochastic nature and dichotomous processing nature, are relatively well aligned with basic aspects of human decision making. Further investigation of the relation of random forests to human decision processes is thus due.

The *Psychological* feature set includes 13 of the building blocks of BEAST, the CPC baseline, and feeding these to a random forest algorithm already produces the best predictive model for the CPC data. Yet, it can be further improved by adding one additional feature: the numeric prediction of the full model, BEAST. That is, in addition to feeding the model with the components of BEAST, it is possible to let it use also its full structure, as was designed by the CPC organizers. This addition implies MSE of 0.0070, relative improvement of 20% on the CPC winner and 29% over BEAST itself.[4] This latter model is also the first model developed for the current data which significantly outperforms the predictions of the baseline BEAST (according to a bootstrap analysis). Therefore, combining a data science tool with the logic derived from psychological theories yields a new state-of-the-art for choice prediction data.

5 Back to Cognition

By integrating psychology and data science, we are able to produce the best predictor for the data. However, many social scientists are interested in predictive models only to the degree that they provide new theoretical insights and/or test existing theories. Our methodology allows such benefits as well.

For example, the theory underlying the baseline model BEAST assumes (put simply) that choice is driven by six behavioral mechanisms. These are (a) sensitivity to the (agents best estimates of the) expected values; (b) minimization of immediate regret and preference for the option better most of the time; (c) neglect of probabilities and treatment of outcomes as equally likely; (d) maximization of probability of gaining and minimization of probability of losing; (e) pessimism (assuming the worst); and (f) special treatment in cases where one option dominates the other. Each of the six mechanisms inspired a subset of the psychological features we derive, as is detailed in Table 4.

The implementation BEAST assumes for the interactions among the six theoretical mechanisms is quite complex, and it is not easy to disentangle each mechanism from the others. Thus, testing the relative importance of each of the six mechanisms is challenging. However, by following the method presented here, a test of the relative importance of these behavioral mechanisms is straightforward. We do this in two ways. First, we simply re-run the random forest algorithm using only those psychological features inspired by five of the six mechanisms and compare its performance to the algorithm using all psychological features. Table 4 shows that running the algorithms without features related to two of the mechanisms, namely sensitivity to the estimated EV and minimization of immediate regret, significantly hurts performance, whereas running the algorithms without features

---

[4]We also tested the other algorithm-feature combinations that include a full BEAST feature and none yields better performance.

Table 3: Test set predictions for algorithm-features combinations

Algorithm	Features used ($MSE * 100$)					
	Obj.	Naïve	Psych.	Obj.+ Naïve	Obj.+ Psych.	Obj.+Naïve +Psych.
Random forest	6.13[*]	1.56[*]	0.98	1.42[*]	0.93	**0.87**
SVM						
radial	5.52[*]	1.63[*]	1.08	1.72[*]	1.10	1.01
polynomial	7.87[*]	5.52[*]	1.23	3.37[*]	1.51[*]	1.40
Neural net (1 hidden)						
3-node	7.39[*]	1.75[*]	1.81[*]	4.80[*]	2.45[*]	2.43[*]
6-node	10.4[*]	2.16[*]	1.89[*]	5.67[*]	2.43[*]	2.40[*]
12-node	10.0[*]	2.98[*]	1.98[*]	5.54[*]	2.57[*]	2.28[*]
Neural net (2 hidden)						
3-3 nodes	8.39[*]	1.91[*]	1.62[*]	4.84[*]	2.48[*]	2.61[*]
6-6 nodes	9.29[*]	3.46[*]	1.85[*]	5.21[*]	2.44[*]	2.36[*]
kNN						
k=1	8.17[*]	3.13[*]	1.87[*]	6.03[*]	3.06[*]	2.73[*]
k=3	7.87[*]	2.22[*]	1.64[*]	4.91[*]	2.75[*]	2.56[*]
k=5	7.15[*]	1.95[*]	1.62	4.72[*]	2.46[*]	2.37[*]

[*] indicates performance significantly different from the baseline-model benchmark (BEAST), according to a bootstrap analysis. Entries (lower is better) are averages of 25 runs. In SVM, predictions were truncated to $[0, 1]$ if necessary.

related to the other mechanisms only mildly affects it. Interestingly, removal of the dominance mechanism improves predictive performance, implying that special treatment of dominant options is misguided given the other theoretical mechanisms.

A second way to examine the relative importance of the six theoretical mechanisms is by using random forests built-in feature-importance analysis tool. Yet, it is known that this tool provides biased measures of feature importance when some of the features are correlated (Strobl et al. 2008), and in our case, the features within each mechanism are highly correlated (e.g., the empirical correlation between dEV_0 and dEV_{FB} is 0.93). Therefore, before using the importance tool, we selected one feature from each mechanism and reevaluated the algorithm. The removal of the correlated features only slightly affected the predictions (MSE of 0.0100 with six features, one for each mechanism, compared to 0.0098 with the full set). The results of the feature importance analysis echo those of the previous method. It suggests that the most important mechanisms are sensitivity to the estimated EV and minimization of regret, whereas the least important mechanism is the special treatment of dominant options.

Therefore, while the theory behind BEAST assumes six behavioral tendencies, both analyses imply a simpler summary of behavior: decision makers are mainly sensitive to the option's expected value and to its probability of providing the better payoff (Erev and Roth 2014). Going back to the insurance company example choosing between two incentive schemes (that have identical EV per violation), it seems that to reduce the number of safety violations, the company should select the scheme that deducts a smaller portion of the discount with high probability (i.e. $0.1 for every violation).

Table 4: The importance of six theoretical mechanisms

Mechanism	Related features	Perf. without ($MSE * 100$)
Estimate EV	dEV_0, dEV_{FB}	1.68
Minimize regret	$pBetter_0, pBetter_{FB}$	1.55
Outcomes equally likely	$dUniEV, pBetter_U$	1.04
Maximize P(Gain)	$dSignEV, pBetter_{S0}, pBetter_{SFB}$	1.11
Pessimism	$dMins, SignMax, RatioMin$	1.04
Dominance	Dom	0.94

6 Discussion

When and how can social scientists and data scientists learn from one another? Currently, members of both communities tend to underestimate the extent to which such learning is possible. We believe this is partly because they tend to focus on different problems. Specifically, many social scientists concentrate on understanding and explaining of behavior, often avoiding making quantitative predictions. Most data scientists, in contrast, focus on tasks for which large amounts of data exists, thus tending to ignore data stemming from controlled laboratory experiments that allow careful examination of human behavior. Our paper tries to address this gap by tackling a prediction problem of choice behavior in controlled experiments.

One major advantage of using this data is the fact that many social science teams made significant attempts to develop models for its prediction, as part of a CPC. This provides us with a strong benchmark to test whether and how proven data-analytic tools can outperform the best social scientists achieve. Interestingly, an improvement over the

best psychological benchmark is difficult to attain. Without psychologically-driven features underlying this benchmark, predictions are significantly worse. Yet, integrating psychological insights with a random forest algorithm does lead to superior performance. Thus, it is possible that the best prospect for development of predictive models of human behavior lies in interactions between social scientists and data scientists. The former would develop theory-grounded features, while the latter would provide the best architecture for their integration.

7 Acknowledgments

References

Allais, M. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society* 21(4):503–546.

Azaria, A.; Rabinovich, Z.; Kraus, S.; Goldman, C. V.; and Gal, Y. 2012a. Strategic advice provision in repeated human-agent interactions. In *Proceedings of AAAI*, 1522–1528.

Azaria, A.; Rabinovich, Z.; Kraus, S.; Goldman, C. V.; and Tsimhoni, O. 2012b. Giving advice to people in path selection problems. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 459–466. International Foundation for Autonomous Agents and Multiagent Systems.

Brandstätter, E.; Gigerenzer, G.; and Hertwig, R. 2006. The priority heuristic: making choices without trade-offs. *Psychological review* 113(2):409–432.

De Melo, C. M.; Gratch, J.; and Carnevale, P. J. 2014. The importance of cognition and affect for artificially intelligent decision makers. In *Proceedings of AAAI*, 336–342.

Edwards, W. 1954. The theory of decision making. *Psychological bulletin* 51(4):380–417.

Erev, I., and Barron, G. 2005. On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review* 112(4):912–931.

Erev, I., and Haruvy, E. 2016. Learning and the economics of small decisions. In Kagel, J. H., and Roth, A. E., eds., *The Handbook of Experimental Economics*. Princeton university press, 2nd edition.

Erev, I., and Roth, A. E. 2014. Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences* 111(Supplement 3):10818–10825.

Erev, I.; Ert, E.; Roth, A. E.; Haruvy, E.; Herzog, S. M.; Hau, R.; Hertwig, R.; Stewart, T.; West, R.; and Lebiere, C. 2010. A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making* 23(1):15–47.

Erev, I.; Ert, E.; and Plonsky, O. 2015. From anomalies to forecasts: A choice prediction competition for decisions under risk and ambiguity. Technical report.

Garner, W. R. 1954. Context effects and the validity of loudness scales. *Journal of experimental psychology* 48(3):218–224.

Gilboa, I., and Schmeidler, D. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153.

Golovin, D.; Krause, A.; and Ray, D. 2010. Near-optimal bayesian active learning with noisy observations. In *Advances in Neural Information Processing Systems*, 766–774.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* 47(2):263–292.

Kahneman, D., and Tversky, A. 1984. Choices, values, and frames. *American psychologist* 39(4):341–350.

Markowitz, H. 1952. The utility of wealth. *The Journal of Political Economy* 60(2):151–158.

Noti, G.; Levi, E.; Kolumbus, Y.; and Danieli, A. 2016. Behavior-based machine-learning: A hybrid approach for predicting human decision making. arXiv:1611.10228.

Payne, J. W. 2005. It is Whether You Win or Lose: The Importance of the Overall Probabilities of Winning or Losing in Risky Choice. *Journal of Risk and Uncertainty* 30(1):5–19.

Prada, R., and Paiva, A. 2009. Teaming up humans with autonomous synthetic characters. *Artificial Intelligence* 173(1):80–103.

Savage, L. J. 1954. *The foundations of statistics*. Oxford, England: John Wiley & Sons.

Srivastava, N.; Vul, E.; and Schrater, P. R. 2014. Magnitude-sensitive preference formation. In *Advances in neural information processing systems*, 1080–1088.

Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; and Zeileis, A. 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9(1):1.

Thorngate, W. 1980. Efficient Decision Heuristics. *Behavioral Science* 25(3):219–225.

Tversky, A., and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4):297–323.

Viscusi, W. K. 1989. Prospective reference theory: Toward an explanation of the paradoxes. *Journal of risk and uncertainty* 2(3):235–263.

Wakker, P. P. 2010. *Prospect theory: For risk and ambiguity*. New York: Cambridge University Press.