Combining data science with social science provides state-of-the-art predictions of human choice behavior

Ori Plonsky, Ido Erev, Tamir Hazan, and Moshe Tennenholtz
Technion- Israel Institute of Technology

Introduction

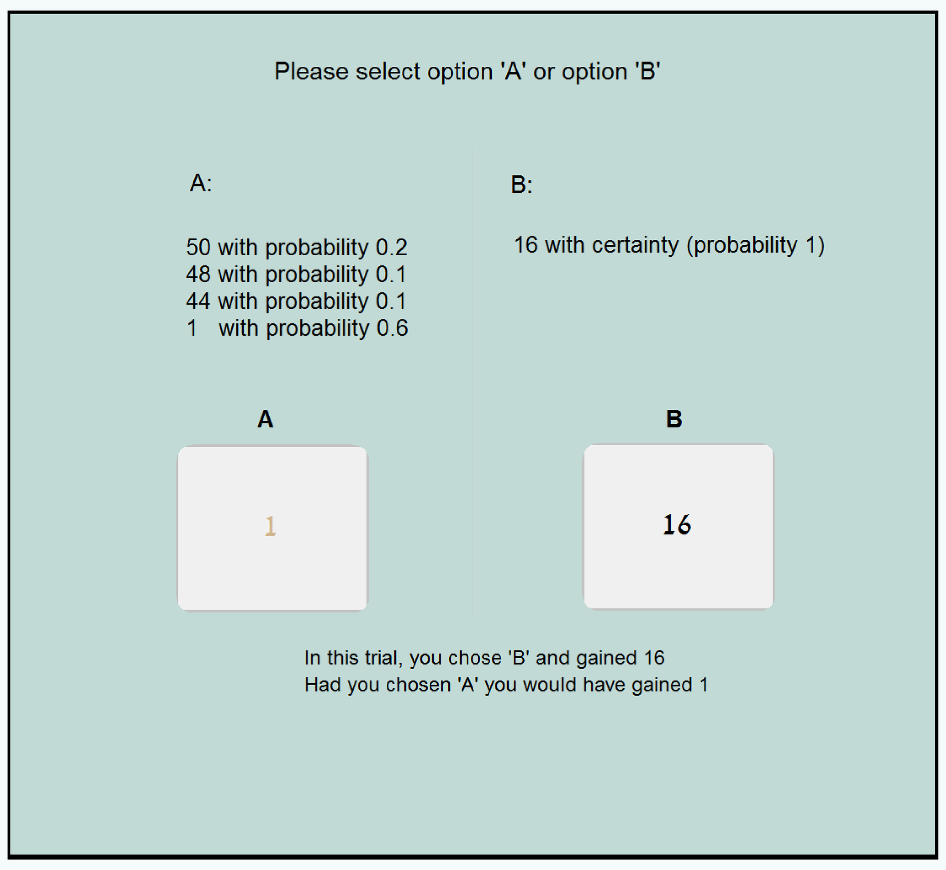
Predicting human behavior is fundamental to both data sciences and behavioral sciences. However, interaction between the two fields is scarce. Here, we introduce a method to integrate the fields and demonstrate its merits in predicting human choice behavior.

In this domain, common practices either (a) focus purely on data scientific tools, mostly neglecting insights from the choice psychology literature; or (b) focus on the psychological drivers of choice, combining them only heuristically, and not rigorously.

We use psychological theories to derive potentially important psychological constructs that we then feed into learning algorithms as features. We thus allow the algorithms to inform us as for the best ways the psychological constructs are to be integrated to a unified predictive model.

Prediction tournament data

To test our method, we use a published dataset (Erev et al., 2017) of consequential human decisions between gambles with full feedback, collected for a models prediction tournament.



Screenshot of the choice experiment from which data is taken.

This data is:

- **Basic**: It comes from a fundamental domain that many agree encapsulates the most basic tenets of choice like attitudes towards risk and value (Kahneman & Tversky, 1979; 1984).
- **Large**: over 330,000 real decisions
- **Challenging**. 25 teams of researchers already developed predictive models for it, as part of the choice prediction tournament.

The idea

Behavior prediction

Social scientists
Develop theory-grounded features.

Data scientists
Provide tools that best integrate the features.

Experiment

To demonstrate the merit of this idea, we examine how well various data science algorithms predict when fed with different types of features.

We define three feature sets:

1. **Objective**: the parameters that define a decision problem (e.g. the outcomes and their probabilities)
2. **Naïve**: basic domain knowledge (e.g. the difference between the options' expected values)
3. **Psychological**: engineered features from useful decision making theories.

For example, decision makers tend to prefer:
(a) the option minimizing probability of immediate regret (Erev & Roth, 2014), hence the feature:

$$pB\_better = P\left[F_B^{-1}(x) > F_A^{-1}(x)\right] - P\left[F_B^{-1}(x) < F_A^{-1}(x)\right]$$

[for the problem on the left, $pB\_better = 0.6 - 0.4 = 0.2$]

(b) the option minimizing the worst possible outcome (Edwards, 1954), hence the feature:

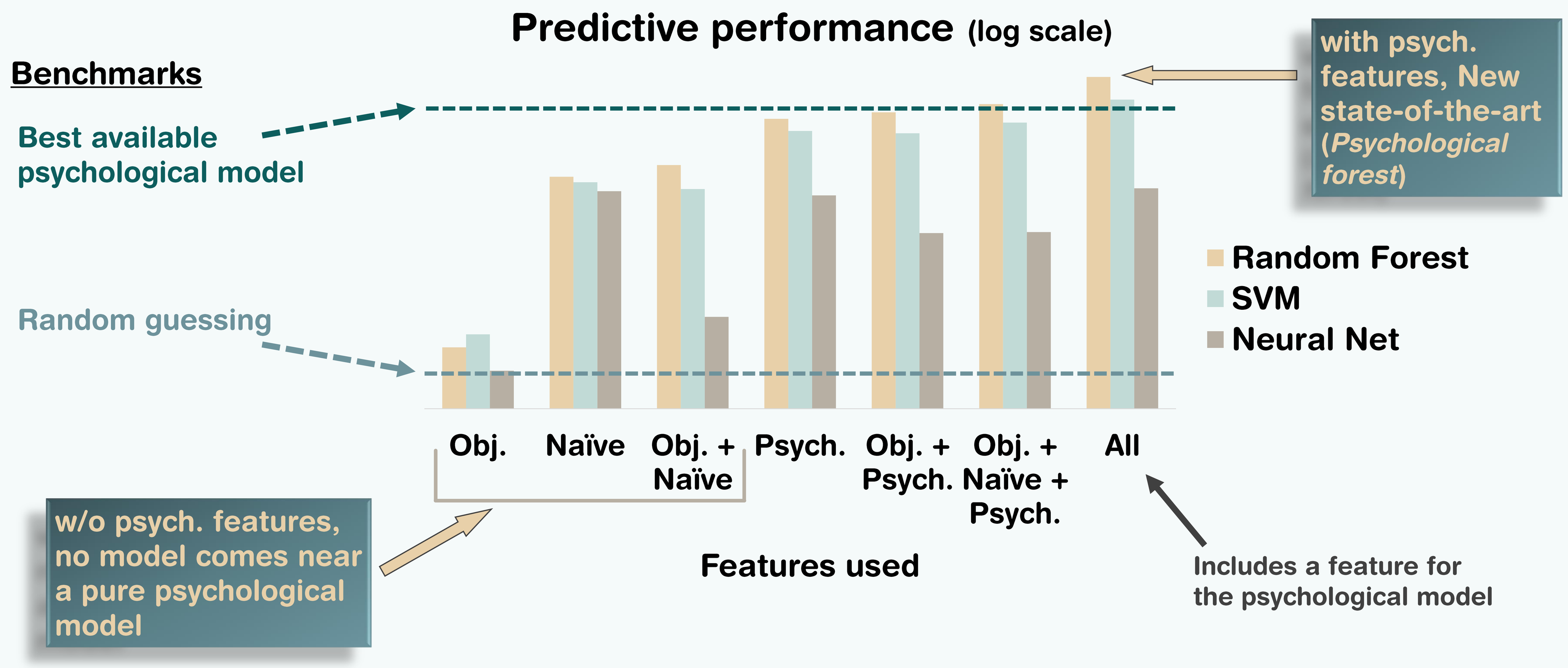
$$diffMins = Min_B - Min_A$$

[for the problem on the left, $diffMins = 16 - 1 = 15$]

(c) the option better assuming all outcomes are equally likely (Thorngate, 1980), hence the feature:

$$dUniEV = EV_{UniB} - EV_{UniA}$$

[for the problem on the left, $dUniEV = 16 - \frac{50+48+44+1}{4} = -19.75$]



Beyond predictions

Pure psychological models normally involve complex interactions between the theoretical constructs they assume. By defining these constructs explicitly as features, we can test the importance of each theoretical element.

For example, the theory behind the psychological model developed for the current data assumes six behavioral tendencies, but using our method, we find a simple summary of behavior: decision makers are mostly sensitive to the option's expected value and to its probability of providing the better payoff.

Discussion

Data scientists and social scientists rarely seek to learn from each other, let alone cooperate.

However, when dealing with behavioral data, data scientists should not overlook the theories developed following years of expert research.

Social scientists, in turn, should not overlook the revolution data science methods brings to the research and application world.

Best practices may emerge from a better collaboration between the fields.

References

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

Erev, I., Ert, E., Plonsky, O., Cohen D., & Cohen O. (2017). From Anomalies to Forecasts: Toward a Descriptive Model of Decisions under Risk, under Ambiguity, and from Experience. *Psychological Review*.

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

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.

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

Acknowledgements

This research was partially supported by the I-CORE program of the Planning and Budgeting Committee and the Israel Science Foundation (grant no. 1821/12).

Paper title:
Psychological Forest: Predicting Human Behavior

Contact Ori Plonsky:
oplonsky@gmail.com