# Modeling Choices at the Individual Level in Decisions from Experience

**Neha Sharma (neha724@gmail.com)**
Applied Cognitive Science Laboratory, Indian Institute of Technology, Mandi, India

**Varun Dutt (varun@iitmandi.ac.in)**
Applied Cognitive Science Laboratory, Indian Institute of Technology, Mandi, India

**Abstract**

Decisions from Experience (DFE) involve situations where decision makers sample information before making a final choice. Trying clothes before choosing a garment and enquiring about jobs before opting for one are some examples involving such situations. In DFE research, conventionally, the final choice that is made after sampling information is aggregated over all participants and problems in a given dataset. However, this aggregation does not explain the individual choices made by participants. In this paper, we test the ability of computational models of aggregate choice to explain choices at the individual level. Top three DFE models of aggregate choices are evaluated on how these models account for individual choices using the maximization criterion. A Primed-Sampler (PS) model, a Natural-Mean Heuristic (NMH) model, and an Instance-Based Learning (IBL) model are calibrated to explain individual choices (maximizing or non-maximizing) in the Technion Prediction Tournament (the largest publically available DFE dataset) and the generalization SC Problems dataset. Our results reveal that all the three DFE models of aggregate choices perform average to explain individual choices. Although the IBL model performs slightly better than PS and NMH models; all the three models are able to account for all individuals in both the calibration and generalization datasets. We conclude by drawing implications for computational cognitive models in explaining individual choices in DFE research.

**Keywords:** Aggregate choice, individual choice, experience, sampling, computational models, maximization.

## Introduction

The steep rise in number of smartphones has given an ample choice to consumers to experiment with (Emarketer, 2014). As a result, a customer now has the privilege of choosing between a wide range of smartphones. To buy the best, one must sample information about the various handsets before making one final choice for one's preferred smartphone. The act of making choices based upon sampled information, however, is not limited to choosing between smartphones rather, it is a very common exercise involving people in different facets of their daily life (choosing clothes, choosing jobs etc.). In fact, information search before a choice constitutes an integral part of Decisions from Experience (DFE) research, where the focus is on explaining human maximizing decisions based upon one's experience with sampled information (Hertwig & Erev, 2009).

In order to study people's search and choice behaviors in the laboratory, DFE research has proposed a "sampling paradigm" (Hertwig & Erev, 2009). In the sampling paradigm, people are presented with two or more options to choose between. These options are represented as blank buttons on a computer screen. People are first asked to sample as many outcomes as they wish from different button options (information search). Once people are satisfied with their sampling of the options, they decide from which option to make a single final choice for real.

Computational cognitive models of human choice behavior have thus far predicted choices at an aggregate level in the sampling paradigm, i.e., when people's final choices are averaged over a large number of participants models (Busemeyer & Wang, 2000; Gonzalez & Dutt, 2012; Lejarraga, Dutt, & Gonzalez, 2012). For example, the Primed-Sampler (PS) model, the Natural-Mean Heuristic (NMH) model, and the Instance-Based Learning (IBL) model are popular DFE algorithms for explaining aggregate choices (Erev et al., 2010; Gonzalez & Dutt, 2011). The PS model depends upon the recency of sampled information, where the model looks back a few samples on each option before making a final choice (Gonzalez & Dutt, 2011). On the other hand, the NMH model is a generic case of the PS model. In this model, one calculates the natural mean of outcomes observed on each sampled option, and using the same for making a final choice (Hertwig, 2011). Similarly, the IBL model (Gonzalez & Dutt, 2011) consists of experiences (called instances) stored in memory. Each instance's activation is a function of the frequency and recency of the corresponding outcomes observed during sampling in different options. These activations are used to calculate the blended values for each option, thereby helping the model make a final choice. The IBL model rely on ACT-R framework for its functioning (Anderson & Lebiere, 1998).

Prior DFE research has shown that, at the aggregate level, the PS, NMH, and IBL models exhibit superior performance compared to other computational models in the sampling paradigm (Erev et al., 2010; Gonzalez & Dutt, 2011). For evaluating these models at the aggregate level, a comparison is made between a model's data and human data from the Technion Prediction Tournament (TPT) dataset (TPT being the largest publically known DFE dataset). However, up to now, none of the three DFE models have been evaluated in their ability to account for maximizing choice behavior at the individual participant level (i.e., in explaining the maximizing final choice of each human participant playing a problem in a dataset). If these models are able to account for maximizing choices at the aggregate level, then one expects that they might also be able to account for maximizing choices at the individual level. However, given that there are sources of noise in both the sampling data as well as in these models, it is likely that these models are no better than random chance in explaining choices at the individual level. Furthermore, as models developed at the aggregate level have different number of free parameters, it is important to use generalization as a test of a model's ability to account for data at the individual level (Busemeyer & Wang, 2000).

In this paper, our main goal is to evaluate how models, which explain choice behavior at the aggregate level, perform at the individual level. In order to evaluate models at the level of an individual participant, we use the largest publically available TPT dataset in the sampling paradigm (Erev et al., 2010) to calibrate the models. Next, we use the SC Problems six-problem dataset (Hertwig et al., 2004) to generalize the calibrated models and test them at the individual level. In what follows, we detail the dataset used and the working of the three models described above. Then, we discuss the methodology of calibrating these models at the individual participant level so as to capture the maximizing final choice. Next, we present the results of models' evaluation at the individual level both during calibration and during generalization. Finally, we close the paper by discussing the implications of our results for models of aggregate choice.

## The Technion Calibration Dataset

The Technion Prediction Tournament (TPT) (Erev et al., 2010) was a competition in which several participants were subjected to an experimental setup, the "e-sampling condition." In this condition, participants sampled the two blank button options in a binary-choice problem before making a final choice for one of the options. During sampling, participants were free to click both button options one-by-one and observe the resulting outcome. Participants were asked to press the "choice stage" key when they felt that they had sampled enough (but not before sampling at least once from each option). The outcome of each sample was determined by the structure of the relevant problem. One option corresponded to a safe choice: Each sample provided a medium (M) outcome. The other option corresponded to the payoff distribution of a risky choice: Each sample provided a High (H) payoff with some probability (pH) or a low (L) payoff with the complementary probability (1 - pH). At the choice stage, participants were asked to select once between the two options. Their choice yielded a random draw of one outcome from the selected option and this outcome was considered at the end of the experiment to determine the final payoff. Competing models submitted to TPT were evaluated following the generalization criterion method (Busemeyer &Wang, 2000), by which models were fitted to choices made by participants in 60 problems (the estimation set) and later tested in a new set of 60 problems (the test set) with the parameters obtained in the estimation set. The 120 problems consisted of choice between a safe option and a risky option as described above. The M, H, pH, and L in a problem were generated randomly, and a selection algorithm was used so that the 60 problems in each set differed in its M, H, pH, and L from other problems. In all the models described here, we have considered an individual human or model participant playing a problem in a dataset (competition or estimation) as an "observation." Also, all model parameters have been calibrated by using the entire TPT dataset that consisted of 120 problems and 2,370 observations. For more details about the TPT, please refer to Erev et al. (2010).

In this section, we detail the working of three popular DFE models that have been used to evaluate human choices at the aggregate level.

### Prime Sampler (PS) Model

The PS model (Hertwig, 2011) employs a simple choice rule. In this model, participants are expected to take a sample of $k$ draws from each option. The exact value of $k$ differs between observations (an observation is defined as a participant playing a problem in a dataset). The PS model assumes that the exact value of $k$ for an observation is uniformly drawn as an integer between 1 and $N$, where $N$ is a free parameter that is calibrated in the model. The final choice for each observation is determined by the following choice probability:

$$Prob\ (Option\ X) = EXP(S_{Mean}X)/(EXP(S_{Mean}X) + EXP(S_{Mean}Y)) \quad \ldots (1)$$

Where, $S_{Mean}X$ and $S_{Mean}Y$ are the samples means of the two options and $Prob\ (Option\ X)$ is the probability of choosing $Option\ X$. For each model observation, the $Prob\ (Option\ X)$ is compared with a random number U (0, 1) to make a choice for one of the two options. If the value of random number is less than or equal to $Prob\ (Option\ X)$, then a choice is made for *Option X*. According to literature, the PS model has performed very accurately at predicting aggregated human choices in the sampling paradigm (Erev, Glozman, & Hertwig, 2008).

### Natural Mean Heuristic (NMH) Model

The NMH model (Hertwig & Pleskac, 2010) involves the following steps:
Step 1. Calculate the natural mean of observed outcomes for each option by summing, separately for each option, all $n$ experienced outcomes and then dividing by $n$.
Step 2. Apply equation 1, where the sample mean is replaced by natural mean.

Thus, the NMH model is a special case of the PS model, where $k$ = an observation's sample size. There are no free parameters in the NMH model. Like the PS model, this model has also performed very accurately at predicting aggregated human choices in the sampling paradigm (Hertwig & Pleskac, 2010).

### Instance Based Learning (IBL) Model

The IBL model is based upon the ACT-R framework (Gonzalez & Dutt, 2011; 2012) and this model is known to predict human aggregate choices better than several DFE models that include assumptions similar to those made in the PS and NMH models (Gonzalez & Dutt, 2011). In this model, every occurrence of an outcome on an option is stored in the form of an *instance* in memory. An instance is made up of the following structure: SDU, here S is the current situation (a number of blank option buttons on a computer screen), D is the decision made in the current situation (choice for one of the option buttons), and U is the goodness (utility) of the made decision (the outcome obtained upon making a choice). When a decision choice needs to be made, instances belonging to each option are retrieved from memory and blended together. Blended values are a function of activation of instances being blended. Activation is a function of the frequency and recency of observed outcomes that occur on choosing options during sampling. In binary choice, the IBL model chooses one of two options by selecting the one having a value greater than a random variable (Gonzalez & Dutt, 2011; 2012). The blended value of option *j* (e.g., a gamble that pays $4 with .8 probability or $0) at any trial *t* is defined as

$$V_{j,t} = \sum_{i=1}^{n} p_{i,t} x_{i,t} \quad \cdots (2)$$

where $x_{i,t}$ is the value of the U part of an instance *i* (e.g., either $4 or $0, in the previous example) at trial *t* and $p_{i,t}$ is the probability of retrieval of that instance from memory at the same trial [10]. Because $x_{i,t}$ is the values of the U part of an instance *I* at trial *t*, the number of terms in the summation changes when new outcomes are observed within an option *j* (and new instances corresponding to observed outcomes are created in memory). Thus, *n*=1 if *j* is a safe option with one possible outcome. If *j* is a risky option with two possible outcomes, then *n*=1 when one of the outcomes has been observed on an option (i.e., one instance is created in memory) and *n*=2 when both outcomes have been observed (i.e., two instances are created in memory).

At any trial *t*, the probability of retrieval of an instance *i* is a function of the activation of that instance relative to the activation of all instances created within that option, given by

$$p_{i,t} = \frac{e^{A_{i,t}/\tau}}{\sum_j e^{A_{j,t}/\tau}} \quad \cdots (3)$$

Where τ, is random noise defined as $= \sigma . \sqrt{2}$ and σ is a free noise parameter. Noise in Equation (2) captures the imprecision of recalling past experiences from memory. The activation of an instance corresponding to an observed outcome in a given trial is a function of the frequency of the outcome's past occurrences and the recency of the outcome's past occurrences (as done in ACT-R). At each trial *t*, activation A of an instance *i* is

$$A_{i,t} = \sigma \ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right) + \ln \sum_{t_i \varepsilon \{1,\ldots,t-1\}}^{n} (t-t_i)^{-d} \quad \ldots (4)$$

where *d* is a free decay parameter; $\gamma_{i,t}$ is a random draw from a uniform distribution bounded between 0 and 1; and $t_i$ is each of the previous trial indexes in which the outcome corresponding to instance *i* was observed. The IBL model has two free parameters that need to be calibrated: *d* and σ. The *d* parameter controls the reliance on recent or distant sampled information. Thus, when *d* is large (> 1.0), then the model gives more weight to recently observed outcomes in computing instance activations compared to when *d* is small (< 1.0). The σ parameter helps to account for the sample-to-sample variability in an instance's activation. For each model observation, the model applies equation 1 to make a choice for one of the two options (for this purpose, the sample mean is replaced by blended values, $V_{j,t}$ for each option).

## The Coin Toss (CT) Model

The CT model is used as a baseline model and it represents chance performance. In this model, we compare the value of a random number between [0, 1] with probability = 0.5. In a binary-choice task, if the random number value < 0.5, then the model chooses the final choice as one option; otherwise, the model chooses a final choice as the other option. When simulated, for a binary-choice task, this model is expected to produce close to 50% accuracy in explaining participants' individual choices. As the probability is fixed at 0.5, this model contains no free parameters.

# Method

## Model Execution

Models submitted to the TPT were evaluated only according to their ability to account for aggregate choice behavior (i.e., the proportion of choices for the option with H and L outcomes were aggregated across participants and problems) (Erev et al., 2010). In this paper, we account for the choice at the individual participant level. For this purpose, a choice made by a model observation is evaluated against a choice made by a corresponding human observation. In order to compare human and model choices for each observation, we evaluate an "error ratio" (i.e., the ratio of incorrectly classified final choices between model and human observations divided by the total number of observations). Firstly, for each observation in human data, we determine the final choice whether maximizing or non-maximizing. In the TPT dataset, a choice is classified as maximizing if the expected value of an option (based upon given problem) with high or low outcome is greater than expected value of an option with medium outcome. Those cases for which the aforementioned criteria fails are termed as having non maximizing choice. Furthermore, human final choice is then compared with the theoretical maximizing or non-maximizing choices to obtain maximizing human final choice per observation. A similar final maximizing choice is then derived for a model observation and this derived choice is compared to the maximizing choice made by the corresponding human observation. The final choices from each of the three models are compared to 2,370 human observations, i.e., the total number of participant-problems-set combinations in TPT dataset. For a model, the error-ratio is calculated as:

Error Ratio = ($M_HN_M$ +$N_HM_M$)/ ($M_HN_M$ + $N_HM_M$ + $N_HN_M$ +$M_HM_M$) ··· (5)

Where, $M_HN_M$ was the number of observations where the human made a maximizing choice but the model predicted a non-maximizing choice. $N_HM_M$ was the number of observations where human made a non-maximizing choice but the model predicted a maximizing choice. Similarly, the $M_HM_M$ and $N_HN_M$ were the number of observations, where a human observation made the same choice (maximizing or non-maximizing) as predicted by the model. The smaller the value of the error ratio, the more accurate is a model in accounting for maximizing individual choices of human participants. For some observations, a model could be equally likely to choose either of the options. Such cases were discarded from the error ratio calculation and were termed as uncategorized (UN) cases, separately. Thus, more are the number of UN cases, the poorer is the corresponding model's algorithm in accounting for complete human data.

In the PS model, an integer value was drawn between 1 and *N*. For the purposes of model calibration, a maximum value of *N* was assumed to be 216. This choice of maximum value was justifiable as 216 is the maximum sample size in the TPT dataset (Erev et al., 2010). For each new value of *N* from 1 to 216, the PS model was run 5 times over the set of 2,370 observations (i.e., for each run, 2,370 observations were used in the model). Error ratio was computed for each of the 5 runs and these five 5 ratios were then averaged for calculating the average error ratio. The *N* value for which the average error ratio was minimized was taken as the calibrated *N* value. The choice of 5 runs of the model is because it enabled us to account for the randomness present in the model due to equation 1.

The NMH model did not possess any free parameters. The model's calibration involved only the natural mean computation for each observation based on the outcomes observed on both the op options during sampling. In NMH model, the final maximizing choice was made for an option based upon equation 1. This model was run 5 times across 2,370 observations since the computation of final choice involved random decisions.

The IBL model described here has two free parameters *d* and σ that were calibrated using a genetic algorithm program. The genetic algorithm repeatedly modified a population of individual parameter tuples in order to find the tuple that minimized the error ratio. In each generation, the genetic algorithm selected individual parameter tuples randomly from a population to become parents and used these parents to select children for the next generation. Over successive generations, the population evolved toward an optimal solution. The population size used here was a set of 20 randomly-selected parameter tuples in a generation (each parameter tuple was a particular value of *d* and σ). The mutation and crossover fractions were set at 0.1 and 0.8, respectively, for an optimization over 150 generations. For each parameter tuple, the IBL model was run 5 times across 2,370 observations. Across the 5 runs, the model's average error ratio was computed by averaging the error ratios from each run. The parameter tuple that minimized the average error ratio across 150 generations were reported as the calibrated parameters for the IBL model.

## Results

In the PS model, the best average error ratio across 5 runs was found to be 0.40. This error ratio occurred at *N*=18. Figure 1 shows the average error ratio results obtained from the PS model for different values of *N*. As shown in the figure, for the first few values of *N*, the error ratio reduced rapidly as the size of *N* increased (up to *N* = 18 samples). However, the error-ratio value saturated to a smaller proportion after increasing *N* further. Thus, there was not much variation in the error ratio from *N*=18 to *N*=216.

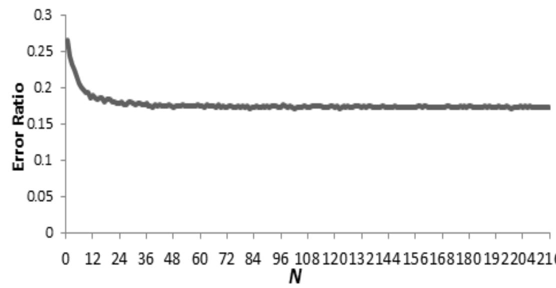


Figure 1. Results from the PS model. The value of parameter *N* was varied between 1 and 216. The best average error ratio = 0.40 for *N*=18.

Table 1 shows the individual-level results from the PS model for the best value of *N*=18. As shown in Table 1, the $N_HN_M$ combinations constituted 24.8% of total combinations. This $N_HN_M$ proportion was second best amongst other combinations ($M_HM_M$, $M_HN_M$, $N_HM_M$, and UN). The $M_HM_M$ combinations had the highest value of 34.3%. The combined average of both the combinations formed about 59.1% of correctly predicted choices by the model. In contrast, the erroneous $M_HN_M$ and $N_HM_M$ ratios were at 21.2% and 19.5% respectively. There were no uncategorized (UN) observations out of 2,370 observations. The PS model's best average error ratio is almost 10% better than the average error ratio of 50% resulting from the CT model. Thus, the PS model explained certain proportion of human dataset fairly accurately.

Table 1. Results from the calibrated PS model. The error ratio = 0.40 for N = 18.

| Combinations from Human Data and Model h/m | Number of Observations | Percentage of 2370 Observations |
|---|---|---|
| $N_HN_M$ | 588.2 | 24.8 |
| $M_HM_M$ | 814 | 34.3 |
| $N_HM_M$ | 463.8 | 19.5 |
| $M_HN_M$ | 504 | 21.2 |
| UN | 0 | 0 |

Next, we investigated the performance of the NMH model. Table 2 shows the results from this model. The model's average error ratio equaled 0.44 across 2,370 observations. In the NMH model, the $N_H N_M$ and $M_H M_M$ values accounted for 22.8% and 32.5% of all 2,370 observations, respectively. The best value of error ratio was for the $M_H M_M$ combinations, where about 32.5 % of the 2,370 human observations were accounted by the model. The erroneous $M_H N_M$ and $N_H M_M$ combinations from the model were 23.08% and 21.5% of the 2,370 human observations, respectively. The second best value accounted by the model was that for the $M_H N_M$ combinations. The error ratio obtained from the NMH model is marginally higher than that obtained from the PS model However, the NMH model too, classified all observations. Thus, like the PS model, the NMH model was able to explain a large dataset. The model showed an improvement of 5% over the CT model and performed efficiently at the individual level.

Table 2. The high level results from the NMH model. The average error ratio = 0.44.

| Combinations from Human Data and Model H/M | Average Number of Observations | Percentage of 2370 Observations |
|---|---|---|
| $N_H N_M$ | 542.4 | 22.8 |
| $M_H M_M$ | 771 | 32.5 |
| $N_H M_M$ | 509.6 | 21.5 |
| $M_H N_M$ | 547 | 23.08 |
| UN | 0 | 0 |

Next, we evaluated the IBL model's ability to account for individual final choices in the TPT dataset. The results from IBL model are presented in Table 3. The best calibrated values of $d$ and σ in the IBL model were found to be 13.6 and 0.22, respectively. The large $d$ value exhibited reliance on recency during sampling resulting in maximization. Also, the small σ value exhibited lesser sample-to-sample variability in instance activations. The calibrated IBL model produced 37.6% of $N_H N_M$ combinations and 26.12% of $M_H M_M$ combinations, respectively. Having a total of comparatively higher values for the $N_H N_M$ and $M_H M_M$ combinations increases the accuracy of the IBL model compared to the NMH and PS models. In contrast, the erroneous $N_H M_M$ and $M_H N_M$ combinations were 17.9% and 18.2% respectively from the IBL model and both these percentages were slightly less than those obtained in the NMH and PS models. Thus, the human choices were predicted more correctly by the model for about 11.1% of total observations. This erroneous classification is about 2% higher than the same erroneous classification from the NMH model. Based upon above statistics, the IBL model's performance was better than the PS and NMH models. Also, the number of uncategorized (UN) cases resulting from the IBL model was zero. Furthermore, as the average error ratio from the IBL model was 0.36, the model shows 14% superior performance compared to the CT model.

Table 3. Results from the calibrated IBL model. The calibrated value of parameters d=13.6 and σ=0.22. The average error ratio= 0.36.

| Choice Combinations from Human Data and Model | Average Number of Observations across 5 Runs | Percentage of 2370 Observations |
|---|---|---|
| $N_H N_M$ | 892.6 | 37.6 |
| $M_H M_M$ | 619.2 | 26.1 |
| $N_H M_M$ | 432.8 | 18.2 |
| $M_H N_M$ | 425.4 | 17.9 |
| UN | 0 | 0 |

Table 7 shows the results summary from the PS, NMH, and IBL models in the calibration TPT dataset. The PS model gave an error ratio of 0.43. The NMH model gave an error ratio of 0.44, which was close to value of error ratio from the PS model. However, the IBL model gave an error ratio of 0.36, which was less than that of the other two models. The NMH model considers the complete sample size of each observation as opposed to the PS model, where the PS model takes into consideration only last few samples of each observation. In this regard, we can conclude that the PS model is more efficient than the NMH model as it uses a much smaller proportion of samples for about the same error ratio compared to the NMH model. The IBL model's error ratio is lower than the other two models; also, the IBL model does not have any UN observations.

Table 4. Summary of results from the three DFE models on TPT dataset.

| Model | Parameters | UN Observations | Error ratio |
|---|---|---|---|
| PS | $N = 18$ | 0 | 0.40 |
| NMH | - | 0 | 0.44 |
| IBL | d=13.6,σ=0.22 | 0 | 0.36 |

As the three models have different number of free parameters, we verified the results from these models upon a generalization to a different dataset (Busemeyer & Wang, 2000). For generalization, we used the SC Problems data set. This dataset has 6 problems, out of which 4 are identical to those in the TPT dataset (one option risky and the other safe) and 2 problems are different from the TPT dataset (both options are risky). Table 5 shows generalization results from the PS, NMH, and IBL models (models were run with the parameters derived in the TPT dataset). The IBL model gave the best error ratio of 0.32. Also, the IBL model did not have any UN observations.

Table 5. Summary of results from the three DFE models (SC Problems dataset).

| Model | UN Observations | Error ratio |
|---|---|---|
| PS | 0 | 0.34 |
| NMH | 0 | 0.37 |
| IBL | 0 | 0.32 |

## Discussions & Conclusion

So far, literature in judgment and decision making had compared models by evaluating their maximizing performance at the aggregate level (Gonzalez & Dutt, 2011; 2012). In such comparisons, the average performance from a model was compared to the average performance from human data (the average was computed across several participants). However, in this paper, we compared a model's maximizing performance at the individual participant level. We used three popular and competing models of aggregate human choice and evaluated their abilities in explaining individual human choices over maximization criterion. Overall, in the TPT dataset, our results reveal that all the three models of aggregate choice performed average at the individual level (error ratio <44%) in both the calibration and generalization datasets. The IBL model's strength is in its ability to account for higher number of maximizing human choices across a large number of observations.

Furthermore, it was found that the two models (PS and NMH) find it easier to explain individual maximizing choices compared to individual non-maximizing choices (there were greater proportion of $M_HM_M$ combinations compared to $N_HN_M$ combinations across the two models). However, the IBL model manages to explain the $N_HN_M$ combinations better than the $M_HM_M$ combinations. Also there were no UN cases reported across the three models.

In addition, the PS and NMH also report a higher proportion of erroneous $M_HN_M$ combinations compared to the erroneous $N_HM_M$ combinations as opposed to the IBL model. Thus, both the models were unable to predict a maximizing final choice for human observations, while the IBL was able to do the same and had a fairly lower error ratio. This finding could also be due to the fact that the one of the options contained two outcomes compared to the single outcome in the other option. Thus, the greater variability experienced in one of the option drove a model to make choices that maximized over a constant value; whereas, the same variability drove humans to maximize upon recency. We generalized the models on another dataset to test this reasoning. In this dataset, again the IBL model performs best thereby proving the stability to the model.

We believe that this paper augurs a beginning of a larger research program that plans to launch an in-depth investigation of the presence of $N_HM_M$ and $M_HN_M$ cases among influential models of experiential choice. As part of future research, we would like to investigate problems where there are multiple options rather than current problems where one options were more of binary kind. Such problems with two options would help us investigate the role of variability in affecting contradictory human and model choices as depicted by the $M_HM_M$ cases. Furthermore, in this paper, we took three competing models of experiential choice; however, as part of future research, we plan to extend this investigation to a larger set of DFE models and other application areas that cover other theoretical ideas.

## Acknowledgments

The authors are grateful to Dr. Ido Erev, Technion - Israel Institute of Technology, and Dr. Ralph Hertwig, Max Planck Institute for Human Development, for providing datasets for this paper. Also, the authors would like to thank Tata Consultancy Services' research fellowship to Neha Sharma for supporting this project.

## References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-2817-6

Busemeyer, J. R., & Wang, Y. (2000).Model comparisons and model selections based on the generalization criterion methodology. *Journal of Mathematical Psychology*, 44, 171–189.

Erev, I., Glozman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk Uncertainty* 36:153–177.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., & Hau, R.(2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Mak*ing, 23, 15–47.

Gonzalez, C.,& Dutt, V. (2012). Refuting data aggregation arguments and how the instance-based learning model stands criticism: A reply to Hills and Hertwig *Psychological Review*, *Vol* 119(4),893-898.

Gonzalez, C., & Dutt, V. (2011). Instance-Based Learning: Integrating Sampling and Repeated Decisions From Experience. *Psychological Review*,118(4), 523-551.

Hertwig, R., & Erev, I.(2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13, 517-523.

Hertwig, R. (2012). The psychology and rationality of decisions from experience. *Synthese*, 187, 269-292.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples*?*. *Cognition*, 115,225–237.

Lebiere, C. (1999). Blending: An ACT–R mechanism for Aggregate retrievals. *Paper presented at the 6th Annual ACT–R Workshop at George Mason University*. Fairfax County, VA.

Lejarraga, T. & Dutt, V. & Gonzalez, C. (2012). Instance-Based Learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*,25: 143-153.

Smartphone Users Worldwide Will Total 1.75 Billion in 2014. (2014, Jan 16). Retrieved from http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536