

Supplementary Material

Detailed description of BEAST

The model BEAST – Best Estimate and Sampling Tools – is a model for decisions under risk, under ambiguity, and from experience (Erev et al., 2017). That is, it is designed to capture choice between fully described gambles, as well as choice when the gambles are ambiguous (the decision maker does not get information concerning the probabilities of some outcomes) and repeated choice with feedback. Specifically, the model assumes that agent i chooses Option A over Option B after r trials with feedback if and only if:

$$[BEV_A(r)_i - BEV_B(r)_i] + [ST_A(r)_i - ST_B(r)_i] + e(r)_i > 0$$

Where $[BEV_A(r)_i - BEV_B(r)_i]$ is the advantage of Option A over Option B based on the *Best Estimates* of their expected values; $[ST_A(r)_i - ST_B(r)_i]$ is the advantage of Option A over Option B based on the output of a correlated mental sampling procedure using *Sampling Tools*; and $e(r)_i$ is an error term. If one option stochastically dominates the other, it is assumed that $e(r)_i = 0$ for all r . Otherwise, this error term is normally distributed with mean 0 and standard deviation σ_i (a property of agent i).

The current paper involves only one-shot choices between fully described gambles and without feedback. Therefore, in the current description of the model, we omit details concerning parts of the model that emerge when the setting is ambiguous and/or when feedback is involved – see Erev et al., 2017 for these details. In our simple setting, the “best estimates” of the expected values (for all agents) are the expected values themselves because they can be computed directly. Hence, the term $[BEV_A(r)_i - BEV_B(r)_i]$ above reduces simply to the difference between the expected values, $EV_A - EV_B$.

The correlated mental sampling process consists of mentally sampling κ_i (property of agent i) outcomes from each option, with each outcome sampled using one of four possible sampling tools: Unbiased, Uniform, Sign, and Contingent Pessimism. In each of the κ_i sampling instances, the agent uses the same sampling tool to draw outcomes from both options. The choice of sampling tool to use in each of the κ_i sampling instances is independent of the other instances (i.e., a specific sampling tool may be used in any

number of instances between 0 and κ_i). In each instance, sampling tool Unbiased is used with probability $1/(\beta_i + 1)$, whereas each of the other sampling tools is used with probability $\beta_i/3(\beta_i + 1)$, where $\beta_i > 0$ is a property of agent i that captures its tendency to use the three “biased” tools.

Sampling of outcomes in all sampling tools uses a “luck level procedure” that implies positive correlation between the sampled outcomes in each sampling instance $j = 1, \dots, \kappa_i$. Under this procedure, to draw an outcome from some distribution F , the agent first draws a luck level, $l_j \sim \text{Uni}(0,1)$, and then draws from F the outcome that corresponds with its current “luck”, $F^{-1}(l_j)$, with F^{-1} the inverse cumulative distribution function. Importantly, in each mental sampling instance, the same luck level is used for generating outcomes from both options. That is, if the agent “feels lucky” – high l_j – then the outcomes mentally drawn from both options will be relatively good, whereas if she “feels unlucky” – low l_j – the outcomes drawn from both options will be relatively bad.

The main differences between the sampling tools involve the distribution F from which outcomes are mentally sampled. In sampling tool Unbiased, the distribution is the objective unbiased distribution that the agent receives as input. Sampling tool Uniform implies a tendency to ignore information on probabilities and give equal weighting to each of the possible outcomes that are described. The distributions from which outcomes are sampled are then uniform distributions over the available outcomes. Sampling tool Sign implies a focus on the payoff sign only, ignoring magnitudes of payoffs. Under this tool, each payoff x in the original distribution is replaced with $R \cdot \text{sign}(x)$, where R is the payoff range in the current choice task. This implies that the output of this tool for each alternative is directly influenced by the context. Finally, sampling tool Contingent Pessimism implies a draw of the worst possible outcome in the original distribution. However, under this tool, pessimism (drawing the worst possible outcomes) is only triggered if two conditions are met: first, the choice task includes at least one positive outcome, and second, the ratio between the original distributions’ worst outcomes is sufficiently large. Formally, the latter condition translates to $RationMins \leq \gamma_i$ (where $0 \leq \gamma_i < 1$ is a property of agent i) with $RationMins$ the ratio between the lower of the two worst outcomes and the higher of the two worst outcomes (and is equal to 0 if they have different signs). When either condition is not met, Contingent Pessimism operates

identically to Uniform. Again, this process implies that the output of the tool is directly influenced by the context.

In total, for the current setting, BEAST includes four properties of individual agents. BEAST's predictions, however, concern a population of agents. The individual parameters are assumed to be drawn from uniform distributions that range between the minimal value possible for that property and the model's free parameters: $\sigma_i \sim \text{Uni}(0, \sigma)$, $\kappa_i \sim \text{Uni}[1, 2, \dots, \kappa]$, $\beta_i \sim \text{Uni}(0, \beta)$, and $\gamma_i \sim \text{Uni}(0, \gamma)$.

Example choice task

Figure S1 shows an example choice task of binary decision under risk, used in the current study. This example is from Erev et al., 2017 dataset and concerns choice between 1 with certainty vs. 20 with probability .05 and 0 otherwise.

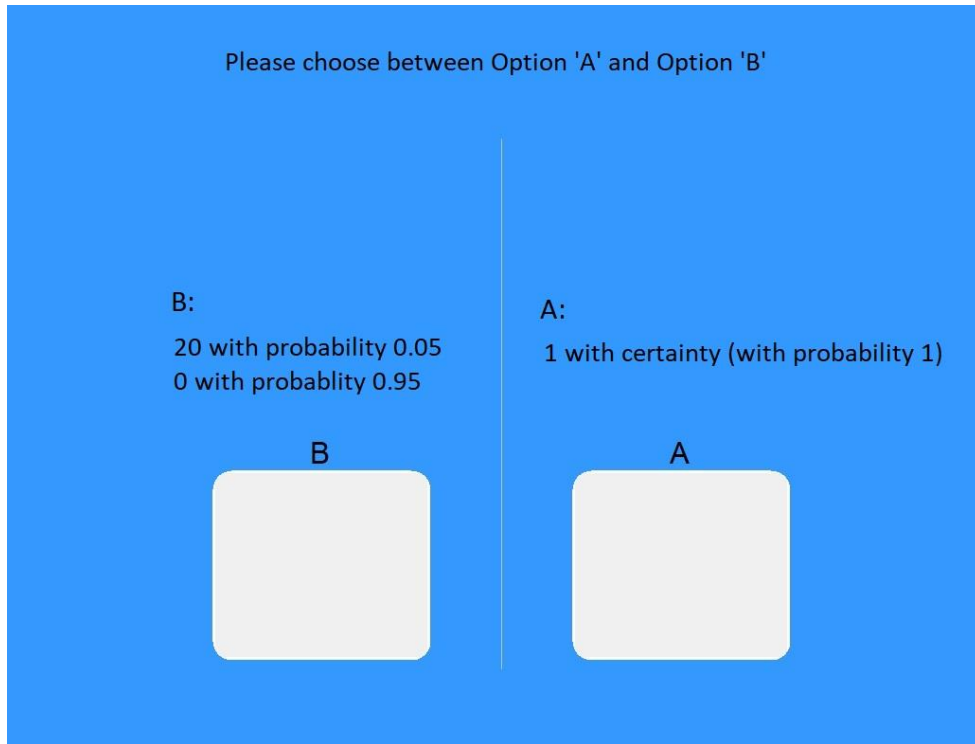


Figure S1. Example task used in the current study. Translated from the original Hebrew and presented in the Erev et al., 2017 dataset.

Estimation Process of the Models

Predictions of both BEAST and AdaBEAST for a choice task require averaging over the output of many simulations of the stochastic process that the models imply, with stable predictions requiring hundreds or thousands of simulations to reduce the noise. This implies that the models do not have analytically specified likelihood functions that can be used for parameter estimation. To estimate the models, we thus used a grid search procedure: for each combination of parameters, we simulated the predictions of the models and examined which parameter combination is the most likely under the data. This procedure, however, requires limiting the set of parameter combinations that are examined (the grid). To determine the values of the grid, for each dataset, we used an iterative process. We initially used a coarse and broad grid search over the parameters, and in each iteration, narrowed down the grid, zooming in on the most likely zones of parameter values in which the models fit best. The values of the parameters that were used in the final iteration for each dataset, that is in the narrowest, most fine-grained grid, are given in Tables S1 and S2. (Unfortunately, we did not save the values of the coarse broader grids we used in earlier iterations for each dataset.) This estimation process was already highly computationally intensive, taking several hours to estimate the parameters for each dataset, hence we did not narrow down the grids further. Note, however, that to the extent that more fine-grained fitting procedure would have results in better parameter values, it is likely that the use of a broader-than-optimal grid only hurt our results. In this sense, the performance we report for our non-optimally fitted non-analytic models may be a lower bound on the true performance possible.

Table S1: Values Used in Estimation of BEAST

Dataset	Parameter's name	Values
Erev17app	β	[0.5, 3.0, 6.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[0.5, 1.0, 2.0, 8.0, 14.0, 18.0, 22.0, 36.0]
	κ	[1, 2, 3, 4, 5, 6, 7]
Fiedler12_exp1	β	[0.5, 3.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 2, 8, 14, 18, 22, 36, 48]
	κ	[1, 2, 3, 4, 5, 6, 7]
Fiedler12_exp2	β	[0.25, 2.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 2, 8, 18, 22, 36, 72]
	κ	[1, 2, 3, 5, 7]
Pachur17	β	[0.25, 2.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 14, 48, 72, 96, 144]
	κ	[1, 2, 3, 4, 5, 6, 7]
Pachur18_e1_session1	β	[0.25, 2.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[2, 8, 18, 27, 36, 50, 72, 96]
	κ	[1, 2, 3, 4, 5, 6, 7]
Pachur18_e1_session2	β	[0.25, 2.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 3, 14, 22, 30, 36, 50, 70]
	κ	[1, 2, 3, 4, 5, 6, 7]
Rieskamp_Positive	β	[0.5, 3.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 2, 14, 18, 22, 36, 48]
	κ	[1, 2, 3, 4, 6, 7]
Stewart15_1A_negative_skew	β	[0.5, 3.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[8, 14, 18, 26, 36, 72, 144]
	κ	[1, 5, 7, 9, 11, 14]
Stewart15_1A_positive_skew	β	[0.25, 2.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 36, 50, 72, 96, 144, 288]
	κ	[1, 2, 3, 7]
Stewart15_1C_positive_skew	β	[0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[0.5, 2.0, 4.0, 18.0, 36.0, 50.0, 72.0, 144.0]
	κ	[1, 2, 3, 4, 7]

Stewart15_2A_negative_skew	β	[0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[0.5, 2.0, 4.0, 8.0, 22.0, 36.0, 50.0, 72.0, 144.0]
	κ	[1, 2, 4, 7]
Stewart15_2A_positive_skew	β	[0.25, 0.5, 3.0, 10.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 4, 8, 18, 36, 72, 108, 144]
	κ	[1, 2, 3, 4, 5, 6, 7]
Stewart15_2B_negative_skew	β	[0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 5.0, 10.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[4, 8, 18, 36, 72, 144, 288]
	κ	[1, 2, 3, 4, 5, 6, 7]
Stewart15_2B_positive_skew	β	[0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 5.0, 10.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[0.5, 1.0, 2.0, 4.0, 8.0, 18.0, 26.0]
	κ	[1, 2, 3, 4, 5, 6, 7]
Stewart16	β	[0.5, 1.5, 3.0, 5.0, 10.0, 15.0, 25.0, 50.0]
	γ	[0.25, 0.5, 0.75, 1.0]
	σ	[1, 4, 36, 72, 144, 288]
	κ	[1, 2]

Table S2: Values Used in Estimation of AdaBEAST

Dataset	Parameter	Values
Erev17app	σ	[0.5, 1.0, 2.0, 8.0, 14.0, 18.0, 22.0]
	p	[0.01, 0.05, 0.1, 0.2, 0.25, 0.3]
Fiedler12_exp1	σ	[0.5, 1.0, 2.0, 4.0, 8.0]
	p	[0.05, 0.1, 0.15, 0.2, 0.25, 0.3]
Fiedler12_exp2	σ	[0.1, 0.2, 0.5, 1.0, 2.0, 8.0]
	P	[0.01, 0.03, 0.05, 0.1, 0.2, 0.25, 0.3]
Pachur17	σ	[1, 2, 8, 14, 18, 27, 36]
	p	[0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4]
Pachur18_e1_session1	σ	[1, 2, 8, 18, 36]
	p	[0.01, 0.03, 0.05, 0.1, 0.2, 0.25, 0.3]
Pachur18_e1_session2	σ	[1, 2, 8, 14, 18, 22, 36]
	p	[0.01, 0.03, 0.05, 0.1, 0.2, 0.25, 0.3]
Rieskamp_Positive	σ	[1, 2, 8, 14, 18]
	p	[0.01, 0.03, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4]
Stewart15_1A_negative_skew	σ	[8, 14, 18, 36, 72, 144]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart15_1A_positive_skew	σ	[0.5, 1.0, 2.0, 4.0, 8.0, 14.0, 18.0, 36.0]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart15_1C_positive_skew	σ	[0.5, 1.0, 2.0, 4.0, 8.0, 14.0, 18.0]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart15_2A_negative_skew	σ	[0.5, 1.0, 2.0, 4.0, 8.0, 14.0, 18.0, 36.0]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart15_2A_positive_skew	σ	[0.5, 1.0, 2.0, 4.0, 8.0, 14.0, 18.0, 36.0]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart15_2B_negative_skew	σ	[4, 8, 18, 36, 72, 144, 288]
	p	[0.01, 0.02, 0.03, 0.05, 0.15]
Stewart15_2B_positive_skew	σ	[1, 4, 8, 36, 72, 1000, 5000, 50000]
	p	[0.01, 0.02, 0.03, 0.05]
Stewart16	σ	[4, 8, 18, 36, 72, 144, 288, 1000, 10000]
	p	[0.01, 0.02, 0.03, 0.05, 0.3]

Note. In addition, in each dataset, the values for the other three free parameters

(W_{uf}, W_s, W_{cp}) were set to be one of [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

Distributions of BEAST Estimated Parameters

We fitted BEAST to each individual 10 times, once for each cross-validation (CV) fold. The distribution of fitted parameters across individuals and CV-folds per dataset are shown in Figure S2. As can be seen, the optimal fit of the parameter β (Figure S2a) in most datasets in the gain domain is often zero. This indicates an effort of the model to compensate for observed deviations from the maximization to an extreme amount, under the original constraint that the difference between EVs receives considerable weight.

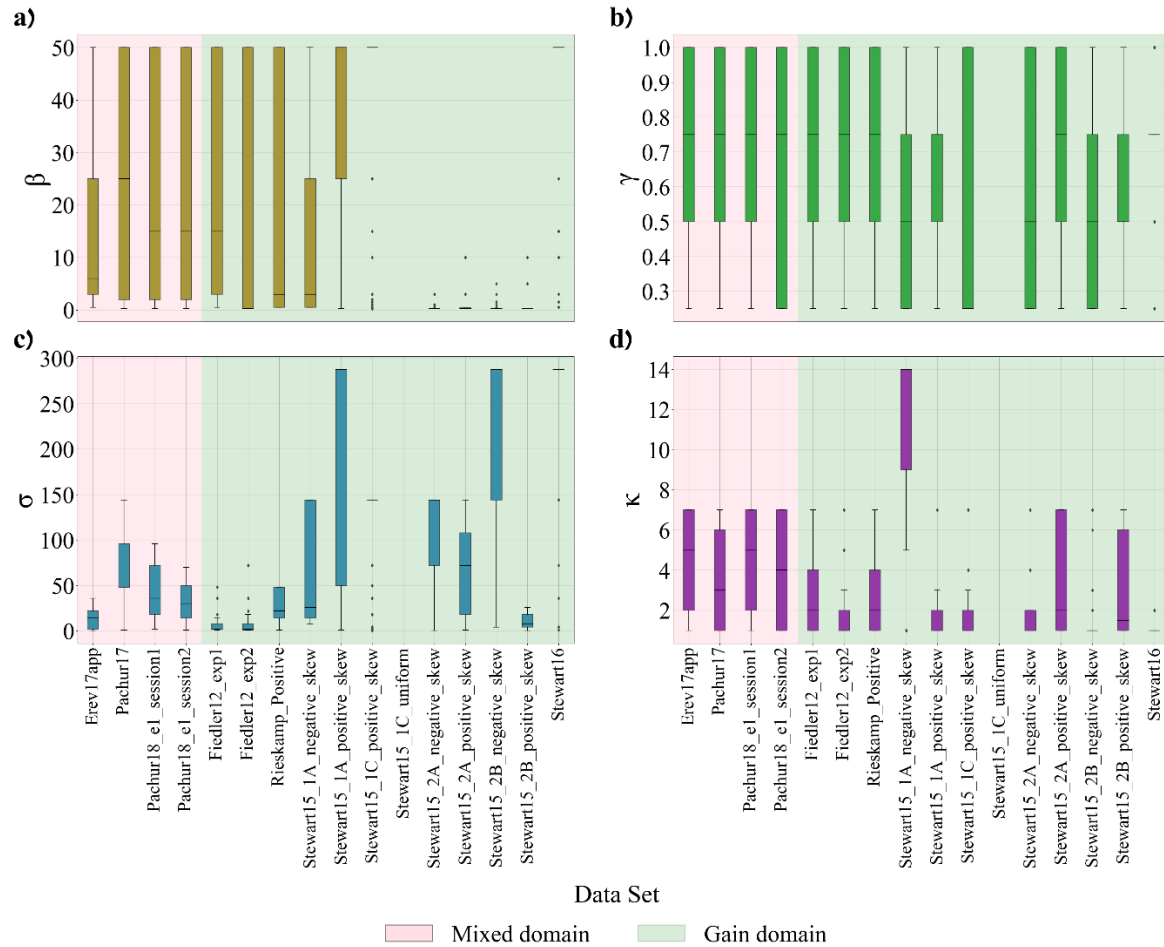


Figure S2. Distributions over participants and CV-folds of BEAST's estimated parameters per dataset. (a) β values showcasing the probability of using the unbiased sampling tool; (b) γ values are used in the lexicographic conditions for the sampling tool *contingent pessimism*; (c) σ values indicating the standard deviation of the error term; and (d) κ values denoting the number of outcomes that are each generated by one of the four sampling tools.

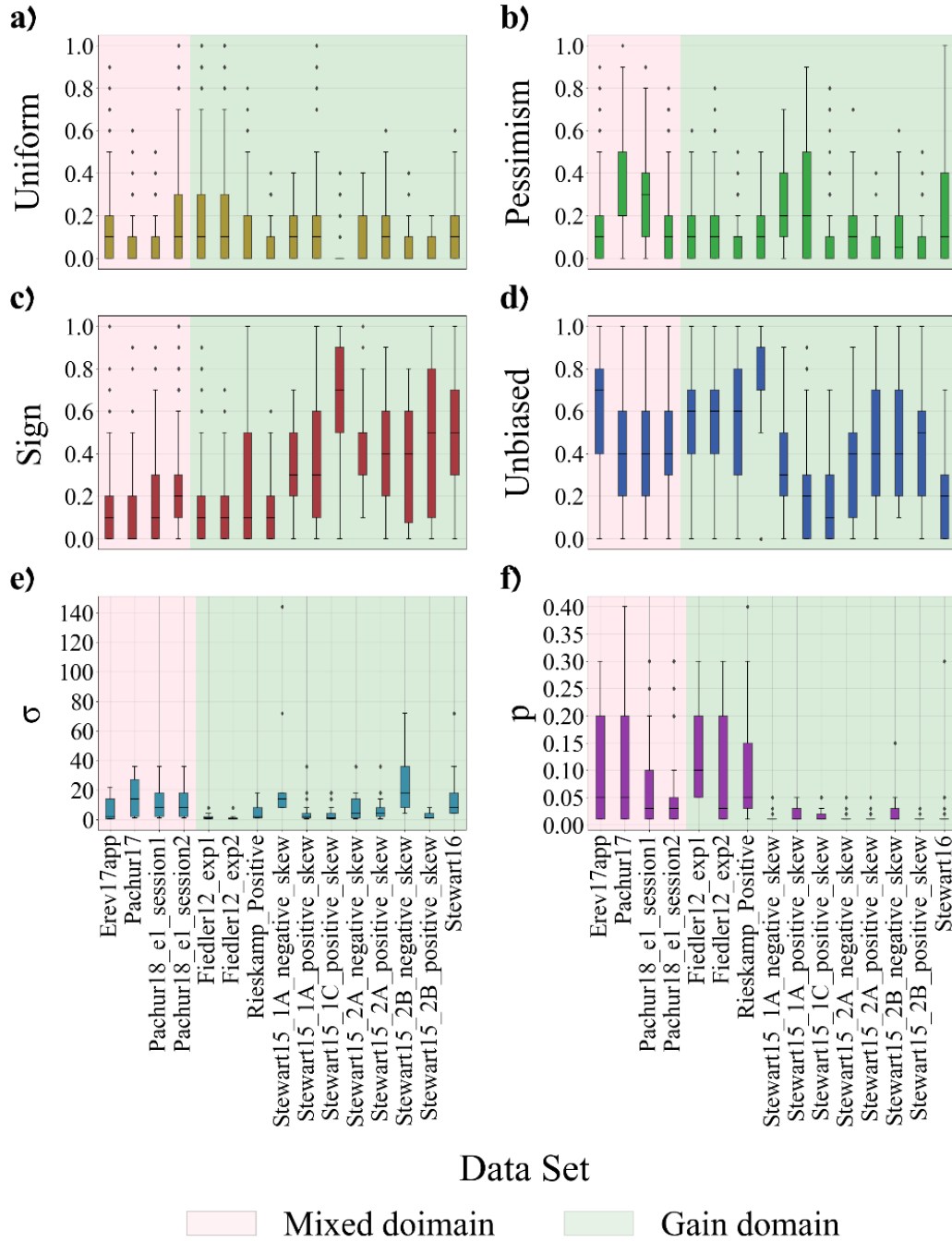


Figure S3: Distributions over participants and CV-folds of AdaBEAST's estimated parameters per dataset. The weights for fitted sampling tools Uniform (a), Pessimism (b) Sign (c) and Unbiased (d); σ values indicating the standard deviation of the error term (e); and p values corresponding to the parameter of a Geometric distribution the sample size is assumed to be drawn from (f).

Usage of Sampling and Regret Mechanisms Among the Models

We repeated the analysis of psychological mechanisms presented in the main analysis, replacing the non-linear payoff and probability transformations mechanisms with the two main mechanisms of BEAST: sampling and regret (Figure S2). The results show that the combination of using both sampling and regret mechanism is unique only for BEAST and AdaBEAST.

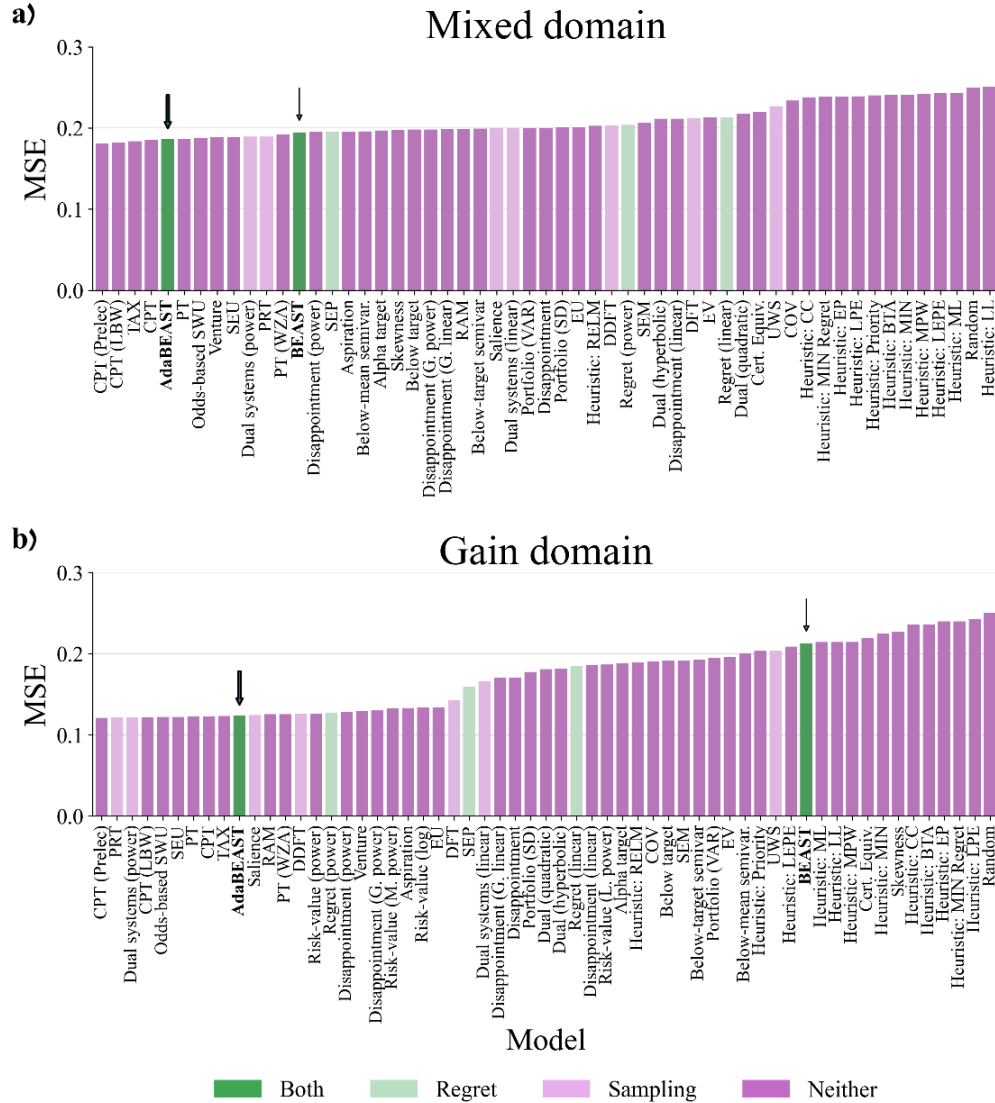


Figure S4: Average prediction error (MSE) for in-sample individuals, for the (a) mixed domain and (b) gain domain. Bar colors indicate usage of sampling and regret by each of the models. Arrows mark the relative ranking of BEAST (thin arrow) and AdaBEAST (thick arrow).

Mixed Effects Statistical Models

Results of a linear mixed effects statistical models with a fixed effect for the behavioral model and random intercepts for participants and for dataset cross-validation fold. Each group in the latter includes one set of problems from one dataset that was predicted by each of the models based on the nine other folds in the same dataset. The statistical models include pairwise comparisons between each two behavioral models' predictions, with negative β values indicating better predictions for the behavioral model listed first in the pair. For convenience, all values (except N) are multiplied by 1000.

Table S3. Regression tables for difference in MSE when predicting familiar individuals' behavior.

	MSE	
	Mixed	Gain
	β (SE)	β (SE)
CPT - BEAST	-13.8 (1.7) ^{***}	-92.1 (2.4) ^{***}
Random Parts (SDs)		
Participant	35.7	51.6
CV fold	18.7	38.9
σ	72.2	93.5
AdaBEAST - BEAST	-8.4 (1.5) ^{***}	-89.2 (2.4) ^{***}
Random Parts (SDs)		
Participant	35	52.4
CV fold	19	38.2
σ	65.6	91.3
CPT - AdaBEAST	-5.4 (1.7) ^{**}	-2.9 (1.7)
Random Parts (SDs)		
Participant	38	44.8
CV fold	19.9	35.4
σ	73.9	65.6
N	7,240	5,920

Note. SE = standard error. CV = cross validation. Number of participants = 658. Number of CV folds = 150.

^{**} $p < .01$, ^{***} $p < .001$.

Table S4. Results for prediction of unknown individuals' behavior.

	MSE	
	Mixed	Gains
	β (SE)	β (SE)
CPT - BEAST	-1.4 (1.4)	
Salience - BEAST		-58.8 (2.2) ^{***}
Random Parts (SDs)		
Participant	38.8	57.3
CV fold	18.4	46.3
σ	61.4	86.1
AdaBEAST - BEAST	-0.2 (1.4)	-56.7 (2.2) ^{***}
Random Parts (SDs)		
Participant	37.4	57.7
CV fold	19	46.7
σ	59.7	85
CPT - AdaBEAST	-1.6 (1.4)	
Salience - AdaBEAST		-2.1 (1.5)
Random Parts (SDs)		
Participant	37.5	52.4
CV fold	18.5	37.9
σ	59.9	56.6
<i>N</i>	7,240	5,920

Note. SE = standard error. CV = cross validation. Number of participants = 658. Number of CV folds = 150.

^{***} $p < .001$.