A combined value formula for risky choice, retired on choices13k

A preregistered test of prospect theory plus information, threat and drive terms against a flexible prospect theory on 11,759 choice problems

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F166 in the findings ledger. Honest null (tier D).

Abstract

We had written a single value formula for risky choice: prospect value plus a term for the information a gamble resolves, a term for the threat of a loss, and a term for drive toward a setpoint. We tested it on the public choices13k data, 1,928 problems answered without outcome feedback and 9,831 answered with it, against cumulative prospect theory and against a flexible prospect theory with the same room to absorb misfit. The kill rule was fixed before any real fit. The combined formula never beat the flexible prospect theory by the preregistered margin: cross-validated R^2 0.634 against 0.645 without feedback and 0.597 against 0.596 with it, against a bar of 0.005. Without feedback, the joint gain of the four added terms, 7.6 in deviance, sat below the 95th percentile of both null controls, 52.0 and 58.9. With feedback two of the terms help as corrections to prospect theory, and the drive term does not. The kill rule fired, and the formula is retired as a model of risky choice. Fitted loss aversion was 1.16 (1.05 to 1.30) without feedback and 0.89 (0.86 to 0.93) with feedback.

1 What this means for a reader

Adding terms to a model of choice will almost always improve its fit a little. The fair test is whether the new terms beat a version of the old theory given the same number of extra parameters. Ours did not, so the terms we named after information, threat and drive describe where prospect theory misfits, and they do not add a new account of why people choose. The loss aversion we measured sits near 1, well below the classic 2.25, which says this number depends on the task that measures it.

2 Introduction

Cumulative prospect theory [2] describes a choice between gambles with three ingredients: a value function that is concave for gains and steeper for losses, a loss aversion coefficient λ that scales losses, and a weighting function that bends probabilities. The theory comes from behavioural economics and describes choices; it gives no account of the feelings behind them. Loss aversion is the factor by which a loss weighs more than an equal gain, so 1 means losses and gains count alike, and it says nothing about a dislike of risk as such. The weighting function maps a stated probability to the weight a choice gives it, so small chances can count for more than their size; it is no belief about the odds. Tversky and Kahneman reported a median λ of 2.25 from their own experiment. Brown, Imai, Vieider and Camerer [3] collected 607 estimates of λ from 150 articles published from 1992 to 2017 and pooled them in a meta-analysis, the statistical pooling of published estimates of one quantity, which collects no new data of its own.

Peterson, Bourgin, Agrawal, Reichman and Griffiths [1] ran the largest risky choice experiment to that date, choices13k, with choice rates on 13,006 problems collected online. They used it to compare classical decision theories with machine-learning models and to search for better theories. Thomas and colleagues [6] argued that the choice behaviour in choices13k contains structured decision noise when compared with a laboratory data set, which bears on how much of the gap between theories and flexible learners is psychology.

The data set also separates problems answered with and without feedback on outcomes. Hertwig, Barron, We-

ber and Erev [5] reported that people choose as if they overweight rare events when choosing from descriptions and underweight them when choosing from experience, the pattern later called the description-experience gap. In decision research the term means that difference between choices from stated odds and choices learned from sampled outcomes; it excludes any difference in the gambles themselves. We analysed the two settings apart for that reason.

Our own starting point was a combined value formula from earlier notes in this project, a name of ours for prospect value plus three added terms in one choice rule, which makes no claim about mechanisms in the brain. It adds to prospect value a term for the information a gamble resolves, a term for the threat of a loss, and a term for reduction of a drive toward a setpoint. This paper is a test of that formula, and the formula is ours to retire.

3 Preregistration and design

The design was written and hash-locked on 31 October 2026 at 08:02 UTC as a preregistration, a practice from clinical trials and psychology: a dated plan fixed before any result is seen, so a plan revised after a look at the outcome does not qualify. Before writing it we had seen the file schemas, the row counts and outcomes simulated by the study code. No real outcome had been fitted.

The locked file names the data, the two settings, every model, the scoring, the planted checks, both controls, the forecasts and the kill rule. A planted check runs data with a known answer through the same analysis first, to show the analysis can find an effect when it is there, which no check on real data can show. A control is a version of the data with the structure under test removed, scored the same way as the real data; a published number from another study is no control. The kill rule decides against the claim when it is met, with no room for judgement after the run. The question it states is whether the combined formula predicts human risky choice better than cumulative prospect theory, and better than a richer pure prospect theory that could absorb the same misfit. Nothing was added after the run. The preregistration's SHA-256 is given in the reproduction section.

Kill and void conditions, as written:

- If the combined formula fails to beat the flexible prospect theory in both settings by 0.005 in cross-validated R^2 with a quasi-binomial BIC change below -10, the formula is retired as a model of risky choice, and its added terms are read as patches for prospect theory misfit whatever their permutation p values say.
- If the planted check fails, every other verdict is void. A verdict, in our usage, is the pass or fail of one preregistered rule set read off by code. A void verdict says nothing for or against the claim, unlike a null result from an instrument that works.

4 Data and provenance

The data are choices13k [1], the selections table and the problems file from the authors' public repository. Each row gives, for one problem, the share of participants who chose option B over option A, the number of participants and the between-person standard deviation of that share. There is no person-level input.

We kept rows marked unambiguous and split them by the feedback flag: 1,928 problems answered without outcome feedback, which we call the description setting because people chose from stated odds alone and learned no outcomes along the way, and 9,831 answered with feedback. In each setting each problem appears once. The repository states no licence. We used a local research copy and report only statistics computed from it.

A join check ran before the lock. The problems file uses row positions in the selections table as its keys, and problem identifiers do not match them. On a dry run with simulated outcomes the join by identifier matched 0.015 of rows; after the fix the match was 1.0 in both settings, checked by recomputing each problem's expected value from both files.

5 Methods

The choice rule

Every value model predicts the log odds of choosing B from the difference in value between the options. The log odds of a rate p is $\ln(p/(1-p))$, a statistics rescaling that is 0 at an even split and adds no information to the rate:

$$\ln\frac{P_B}{1-P_B} = b_0 + \tau(V_B - V_A) \tag{1}$$

where P_B is the predicted rate, b_0 an intercept and τ a sensitivity. Models were fitted by binomial likelihood on the observed rates, the statistics method that picks the parameters under which the observed counts of B choices are most probable, treating each choice as an independent coin with the predicted rate, counting five choices per listed participant, the count the study code uses for each problem.

The models

The expected value baseline (EV), the probability-weighted average payoff from probability theory with no allowance for risk or loss, uses $V = \sum_i p_i x_i$ for outcomes x_i with probabilities p_i. Cumulative prospect theory (CPT) uses rank-dependent weights π_i, decision weights that depend on where an outcome ranks among the gamble's outcomes and are taken as differences of the weighting function at cumulative probabilities, which excludes weighting each probability on its own, and a value function that for gains is

$$v(x) = x^{\alpha} \tag{2}$$

and for losses is $v(x) = -\lambda(-x)^{\alpha}$, with $V = \sum_i \pi_i v(x_i)$. The weights come from the Tversky and Kahneman weighting function applied to cumulative probabilities, with its own γ for gains and for losses:

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \tag{3}$$

CPT has six parameters. The flexible prospect theory (FLEX) has separate curvature for gains and losses, λ, and the two-parameter Prelec weighting function [4] on each side, nine parameters in all. Prelec's function, from decision theory, sets the curvature and the crossing point of the weighting curve separately, a shape the one-parameter function cannot take:

$$w(p) = e^{-\delta(-\ln p)^{\gamma}} \tag{4}$$

The combined formula

The combined formula (COMB) keeps CPT and adds four terms to the log odds of eq. 1, each the value for option B minus the value for option A:

$$\ln\frac{P_B}{1-P_B} = b_0 + \tau(V_B - V_A) + \beta_1(H_B - H_A) + \beta_2(T_B - T_A) + \beta_3(M_B - M_A) + \beta_4(Q_B - Q_A) \tag{5}$$

Here H is the Shannon entropy of the outcome distribution in bits, read as the information gained when the gamble resolves. In information theory entropy measures how uncertain an outcome is before it happens, 0 for a sure thing; it ignores the sizes of the payoffs:

$$H = -\sum_i p_i \log_2 p_i \tag{6}$$

T is the probability of a loss, read as threat. M is the mean outcome and Q the mean squared outcome. Those two carry the drive term: for a setpoint S, the drive reduction of an outcome is

$$S^2 - (S - x)^2 = 2Sx - x^2 \tag{7}$$

so its expectation is linear in M and Q. Drive reduction is a term from motivation psychology, here only the quadratic form above, and it excludes any measured need or state of the people choosing. The mean was divided by 10 and the mean square by 1,000 before fitting. COMB has ten parameters. Three reduced models add each term alone: CPT+H (entropy), CPT+T (loss probability) and CPT+D (mean and mean square).

Scoring

Each model was scored by 10-fold cross-validated R^2 on the rates, with folds grouped by problem. Cross-validated R^2, from statistics and machine learning, is the share of variance in the observed rates that predictions made without seeing those rates account for, so it does not reward fitting noise as in-sample R^2 does. With out-of-fold predictions q_i, observed rates y_i and their mean m_y,

$$R^2 = 1 - \frac{\sum_i (y_i - q_i)^2}{\sum_i (y_i - m_y)^2} \tag{8}$$

Each model was also scored by a quasi-binomial BIC, with the dispersion φ taken from the COMB fit in each setting (5.70 without feedback, 7.58 with it):

$$B_q = -\frac{2 \ln L}{\varphi} + k \ln n \tag{9}$$

where L is the binomial likelihood, k the number of parameters and n the number of problems. The Bayesian information criterion, a statistics score, trades fit against the number of parameters, lower being better; it compares models on the same data and gives no absolute test. The dispersion scales the likelihood for rates that scatter more than independent choices would allow, the quasi-binomial adjustment, and it is no parameter of choice.

Two references bound the value models. Gradient-boosted trees, fitted to the raw problem parameters and the four added terms with the same folds, give a flexible learner's score. The method, from machine learning, adds many small decision trees each fitted to what the previous ones missed; it predicts well and states no theory of choice. A sampling ceiling, our bound, is the share of variance across problems not explained by sampling noise, $C = 1 - m_s / v_y$. Here m_s is the mean over problems of the squared between-person standard deviation divided by the number of participants, and v_y is the variance of the observed rates.

Tests on the added terms

Each added term was tested with the fitted CPT log odds held as a fixed offset: the term was standardised, fitted by iteratively reweighted least squares, and its deviance gain was compared with 200 fits on the same term permuted across problems. An offset, in regression, is a part of the prediction fixed in advance and never refitted, so the term can only explain what CPT leaves over. Iteratively reweighted least squares is the standard fitting routine for such models, a means of fitting with no effect on what is tested. Deviance is twice the log-likelihood shortfall of a model against a perfect fit, so a deviance gain is the improvement in fit from the term, before any penalty for its parameter. The p value is a p value from draws, as in permutation statistics, and assumes no textbook distribution; it counts one added to the numerator and the denominator:

$$p = \frac{k + 1}{n + 1} \tag{10}$$

where k is the number of permuted gains at or above the real one and n the number of permutations. The smallest possible value at 200 permutations is 0.005. The six p values, three terms in two settings, were corrected by Benjamini and Hochberg's procedure at q 0.05. That multiple-testing rule keeps the expected share of false calls among the tests called significant at or below q, which is looser than the Bonferroni bound on any false call; the corrected values are q values.

The loss aversion interval comes from refitting CPT on problems resampled with replacement: 100 resamples without feedback and 30 with feedback, with the 2.5th and 97.5th percentiles reported. This is a bootstrap, the statistics method that reads a number's spread from resampled data with no distribution assumed, and the

interval holds the middle 95 percent of the resamples; it is no statement that the true value lies inside with 95 percent probability.

6 Controls and planted checks

The planted check ran on the description problems with their real participant counts. Three data sets were simulated from CPT with λ 1.6, and three from CPT plus an entropy term with coefficient 0.25, each with extra-binomial noise drawn from a Beta distribution. That is scatter in the choice rate from problem to problem beyond what independent coin flips would give, the kind of spread real aggregate rates show, and it adds no bias to the rate. Selection between CPT and COMB was by the quasi-binomial BIC. The rule required the right model in at least 5 of 6 data sets and λ recovered within 0.3 in at least 2 of 3. Model selection picked the true model in 6 of 6 planted data sets and recovered loss aversion within 0.3 in 3 of 3. The three recovered values were 1.64, 1.55 and 1.59. The check passed, and no verdict was voided.

Two null controls ask whether the description setting's joint gain from the four added terms is larger than chance would give. Their draws form a null distribution, drawn from these data and never assumed from theory, whose 95th percentile is the bar the real gain must clear. The first permutes the four terms together across problems with CPT held fixed (200 draws). The second simulates outcomes from the fitted CPT at the observed dispersion and fits the four terms to those (100 draws). The second is a false-success check: on data with no added effect, it shows how large a gain the procedure finds anyway.

7 Results

Table 1 gives every model's score in both settings, and Figure 1 draws the main rows. Prospect theory improves on expected value by a wide margin in both settings. The flexible prospect theory improves on CPT by 0.017 without feedback and 0.012 with it.

Table 1. Cross-validated R^2 and quasi-binomial BIC by model and setting. Lower BIC is better. k is the number of parameters.

Model	k	R^2, no feedback	BIC, no feedback	R^2, feedback	BIC, feedback
EV	2	0.373	4,870.3	0.464	19,465.6
CPT	6	0.628	3,637.8	0.584	16,745.2
FLEX	9	0.645	3,569.5	0.596	16,479.2
CPT+H	7	0.628	3,645.2	0.587	16,681.1
CPT+T	7	0.628	3,644.9	0.585	16,724.6
CPT+D	8	0.634	3,615.9	0.593	16,527.3
COMB	10	0.634	3,628.1	0.597	16,440.8
Boosted trees	-	0.735	-	0.805	-
Sampling ceiling	-	0.825	-	0.863	-

The combined formula never beat the flexible prospect theory by the preregistered margin: cross-validated R^2 0.634 against 0.645 without feedback and 0.597 against 0.596 with it, against a bar of 0.005. The differences are -0.011 and +0.001. Without feedback COMB also loses on BIC, by 58.7. With feedback COMB has the lower BIC by 38.4, which meets the BIC half of the rule, and the R^2 half fails. The kill rule fired in both settings.

Against basic CPT the picture differs by setting. Without feedback COMB gains 0.006 in R^2, which meets the 0.005 bar, but its BIC change is -9.6, short of -10, and no added term survives correction. With feedback COMB gains 0.013 in R^2 and lowers BIC by 304.4, and that verdict passes. Most of COMB's gain without feedback comes from the drive term alone: CPT+D scores 0.634, the same as COMB to three places.

Without feedback, the joint gain of the four added terms, 7.6 in deviance, sat below the 95th percentile of both null controls, 52.0 and 58.9. Figure 2 shows the real gain at the low end of both null ranges.

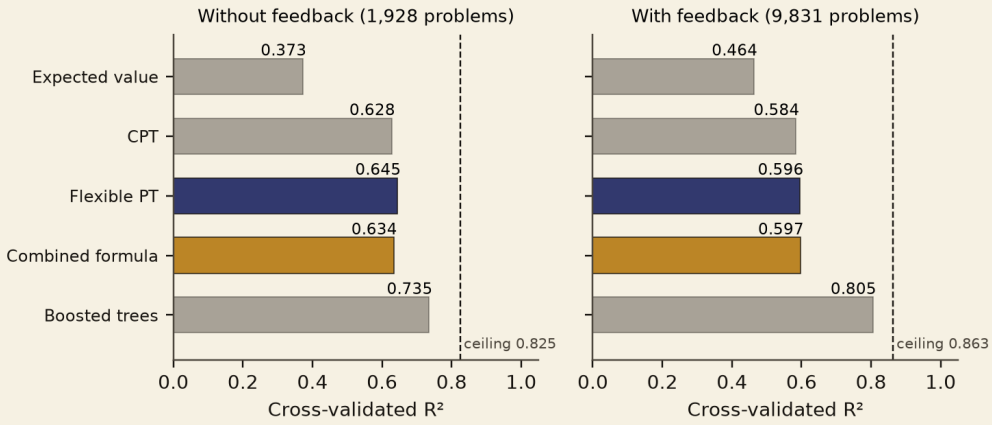


Figure 1. Cross-validated R^2 by model, without and with feedback. The flexible prospect theory (dark) and the combined formula (ochre) have about the same score in both settings. Dashed lines mark the sampling ceiling. Drawn from the run's result file.

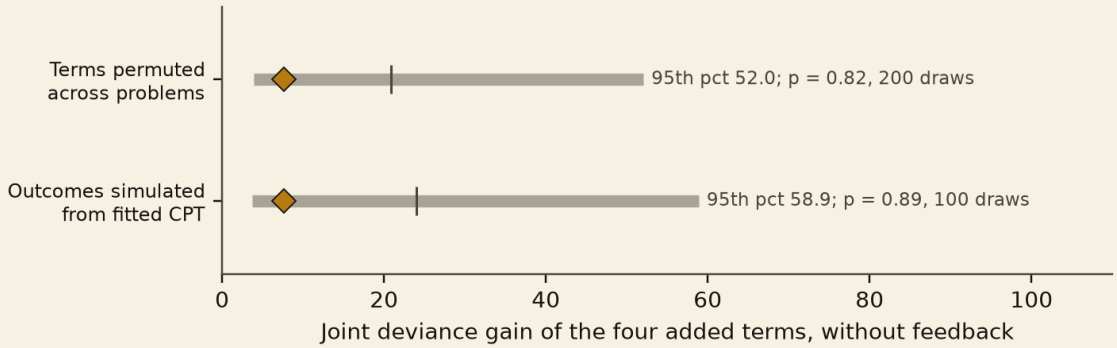


Figure 2. Joint deviance gain of the four added terms without feedback (diamonds) against the 5th to 95th percentile range (bars) and mean (ticks) of each null control.

Table 2 gives the tests on single terms. With feedback, entropy and loss probability each cleared the correction as residual terms on top of prospect theory (q 0.015), and the drive term did not (q 0.46). Both surviving terms sit at the smallest p value 200 permutations can give.

Table 2. Deviance gain of each added term with CPT held as an offset, its permutation p value (200 permutations) and the Benjamini-Hochberg q value over all six tests.

Setting and term	Deviance gain	p	q
No feedback, entropy	0.68	0.77	0.77
No feedback, loss probability	0.48	0.75	0.77
No feedback, drive	6.38	0.55	0.77
Feedback, entropy	475.0	0.005	0.015
Feedback, loss probability	59.5	0.005	0.015
Feedback, drive	25.5	0.23	0.46

Fitted loss aversion was 1.16 (1.05 to 1.30) without feedback and 0.89 (0.86 to 0.93) with feedback. Figure 3 sets both against 1 and against 2.25. The other CPT parameters without feedback were curvature 0.75 and weighting γ 0.78 for gains and 0.57 for losses; with feedback, 0.79, 0.87 and 0.82.

Table 3 sets the forecasts written before the run against the outcomes.

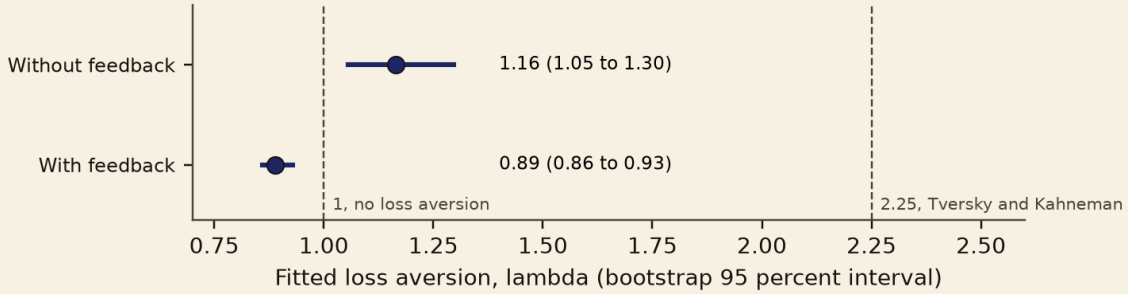


Figure 3. Fitted loss aversion with bootstrap 95 percent intervals: 100 resamples without feedback, 30 with feedback.

Table 3. Forecasts fixed in the preregistration against what the run gave. R^2 is cross-validated.

Quantity	Forecast	Outcome
EV R^2, no feedback	0.35	0.373
CPT R^2, no feedback	0.60 (0.45 to 0.72)	0.628
FLEX minus CPT, no feedback	+0.02	+0.017
COMB minus CPT, no feedback	+0.015	+0.006
COMB minus FLEX, no feedback	+0.005	-0.011
Boosted trees R^2, no feedback	0.75	0.735
Sampling ceiling, no feedback	0.90	0.825
Loss aversion, no feedback	1.3 (1.0 to 1.8)	1.16
CPT R^2, feedback	0.50	0.584
COMB minus FLEX, feedback	+0.01	+0.001
COMB beats CPT, no feedback	probability 0.55	no
COMB beats FLEX, no feedback	probability 0.25	no
Loss aversion interval below 2.25	probability 0.8	yes

8 Discussion

The test was built so that the combined formula had to beat a pure prospect theory with comparable freedom. It did not. Without feedback the added terms carry no signal beyond what the nulls produce, and three extra prospect theory parameters do more than four added terms. With feedback the added terms and the flexible theory tie. The preregistered reading follows: the terms are patches for places where CPT misfits.

The terms themselves suggest which places. The drive term expands into the mean and the mean square of outcomes, so it adds an expected value and a variance correction to CPT. That is a change to the shape of the value function, which FLEX can also make through separate curvatures. Entropy and loss probability help only with feedback. Rare events are underweighted when people learn from experience [5], and a gamble with more outcomes, or with a small chance of loss, is the kind of gamble where that matters. This reading is inferred and was not tested here; a prospect theory with weighting fitted to experience could absorb these terms as well.

Both prospect theories fall well short of boosted trees, by 0.09 without feedback and 0.21 with it, and the trees fall short of the sampling ceiling. Peterson and colleagues [1] reported a similar gap between classical theories and learned models on the same data. Part of the gap may be structure that a value model cannot express, such as correlation between payoffs; part may be the structured noise Thomas and colleagues [6] described. We did not separate the two.

Loss aversion of 1.16 without feedback, with an interval that excludes 1 and lies far below 2.25, adds one more estimate to the spread Brown and colleagues [3] pooled. With feedback the estimate falls below 1. This is an aggregate estimate on small online stakes and says nothing about any single person's loss aversion.

9 Limitations

- Entropy, loss probability and the mean and mean square of outcomes are one reading of information, threat and drive. Another reading could score differently, so the result covers this operationalisation only.
- The data are aggregate rates per problem, so no individual parameters were estimated, and the planned mapping of individual parameter tails was not possible here.
- Stakes are small and the sample was collected online.
- The repository states no licence, so the data are used as a local research copy.
- Each cross-validation fold was refitted from the full-data optimum with a capped number of iterations. That start leaks a little information from the held-out fold, the same for every value model.
- The interval for loss aversion with feedback rests on 30 bootstrap resamples, and the simulated null control on 100 draws.
- The planted check ran on the description problems only.
- One author wrote the forecasts, the rules and the scoring code, and scoring was not blind.

10 Verified

Verified: on 9 November 2026 the problem counts 1,928 and 9,831, every cross-validated R^2 and BIC in Table 1, the dispersions, the differences between models, the deviance gains, p values and q values in Table 2, the joint gain 7.6 and both null controls' means, percentiles and draw counts, the CPT parameters, both loss aversion estimates and intervals, the planted results and all seven verdicts were re-read from the run's result file. The forecasts were re-read from the preregistration. The preregistration's SHA-256 matches the hash recorded in the result file and the lock file. All three figures were drawn by a script that reads only that result file. Methods were checked against the study code, whose hash the result file records. The study was not re-run for this paper. The boosted-trees score without feedback is 0.7355 in the result file, which rounds to 0.735; the ledger entry and an earlier write-up give 0.736, and this paper uses the result file. The findings ledger is our append-only, dated list of every recorded result, where corrections enter as new lines and old entries are never rewritten. The literature in the introduction was found by searches run on 9 November 2026 and read at the level of abstracts, publisher pages and index records; the full texts of the cited papers were not read.

11 Corrections

- 31 October 2026: before the lock, a dry run found that the problems file uses row positions in the selections table as its keys. The join was fixed and a join check added to the planted rule.
- 9 November 2026: first published version. The boosted-trees score without feedback is reported as 0.735, from the result file, in place of 0.736 in the ledger entry.

12 Reproduction

- Data: download choices13k from the authors' public repository (the selections table and the problems file). Join the problems file by row position in the selections table.
- Keep rows marked unambiguous and split them by the feedback flag.
- Fit EV, CPT, FLEX, CPT+H, CPT+T, CPT+D and COMB under eq. 1 by binomial likelihood, with the terms of eqs. 5 to 7.
- Score by 10-fold cross-validated R^2, eq. 8, and quasi-binomial BIC, eq. 9, with the dispersion from COMB. Fit boosted trees on the same folds and compute the sampling ceiling.
- Test each added term with CPT as an offset against 200 permutations, correct the six p values by Benjamini and Hochberg, and run both null controls on the description setting's joint gain.

- Bootstrap CPT's loss aversion over problems: 100 resamples without feedback, 30 with feedback.
- Preregistration SHA-256: cecb9ceb76ab37020c14039391fd85eece47333c4b75a2cf3f87cd0a6133103a. Plan SHA-256: b8e125eabcc4160e6dad5de3d90207cbae261b9a0779fb1f56dd6ded054fba32. Run 20261031T080226Z, seed 20261031.
- Cost: the analysis took 641 seconds on one workstation, and both controls together under one second.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run. A verdict passes only when every rule holds. Values are read from the run's result file.

Table 4. Preregistered verdicts and the rules that decided each one.

Verdict	Deciding rules	Values	Passes
Planted check	selection right above 0.8 of planted sets; loss aversion recovered above 0.6 of sets; join match above 0.999	1.0; 1.0; 1.0	yes
COMB beats CPT, no feedback	R^2 gain at least 0.005; BIC change below -10; smallest q below 0.05; joint gain above both null 95th percentiles (52.0, 58.9)	0.006; -9.6; 0.77; 7.6	no
COMB beats FLEX, no feedback	R^2 gain at least 0.005; BIC change below -10	-0.011; +58.7	no
COMB beats CPT, feedback	R^2 gain at least 0.005; BIC change below -10; smallest q below 0.05	0.013; -304.4; 0.015	yes
COMB beats FLEX, feedback	R^2 gain at least 0.005; BIC change below -10	+0.001; -38.4	no
Loss aversion above 1	lower bound of the interval without feedback above 1	1.05	yes
Loss aversion below 2.25	upper bound of the interval without feedback below 2.25	1.30	yes

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **cumulative prospect theory (CPT)**. A model of choice between gambles with a value function concave for gains and steeper for losses, a loss aversion coefficient and a probability weighting function applied to cumulative probabilities; here six parameters. Field: behavioural economics. Excludes any account of the feelings or brain processes behind a choice.
- **loss aversion**. The factor lambda by which a loss weighs more than an equal gain in the value function; 1 means losses and gains count alike. Field: behavioural economics. Excludes a dislike of risk as such, which curvature and weighting carry.
- **probability weighting function**. A function that maps a stated probability to the weight a choice gives it, so that small chances can count for more than their size and large ones for less. Field: behavioural economics. Excludes a person's belief about the odds, which in these problems are stated.
- **meta-analysis**. The statistical pooling of published estimates of one quantity into a summary and a measure of their spread. Field: statistics, research synthesis. Excludes collecting new data.
- **description-experience gap**. The pattern that people choose as if they overweight rare events when choosing from stated odds and underweight them when they learn the odds from sampled outcomes. Field: decision research. Excludes differences in the gambles themselves.

- **combined value formula (COMB)**. CPT's value difference plus four added terms in one choice rule: entropy for information, loss probability for threat, and the mean and mean square of outcomes for drive toward a setpoint; ten parameters. Field: our usage. Excludes any claim about mechanisms in the brain, and any reading of the terms other than this operationalisation.
- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (metascience). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons found after the run for setting a result aside.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.
- **void**. The verdict of a run whose planted check failed or whose sample fell below its stated minimum; it says nothing for or against the claim. Field: research method (our usage). Excludes a null result, which comes from a working instrument.
- **description setting**. The 1,928 choices13k problems answered without outcome feedback, where people chose from stated odds alone. Field: our usage, after decision research. Excludes the 9,831 problems answered with feedback on outcomes.
- **log odds**. For a rate p, the natural log of p divided by 1 minus p; 0 at an even split and unbounded toward certainty. Field: statistics. Excludes any information beyond the rate, of which it is a rescaling.
- **binomial likelihood**. Fitting by choosing the parameters under which the observed counts of B choices are most probable, treating each choice as an independent draw at the predicted rate; here five choices per listed participant. Field: statistics. Excludes extra scatter between problems, which the dispersion handles.
- **expected value (EV)**. The probability-weighted average payoff of a gamble, used here as the baseline value model. Field: probability theory. Excludes any allowance for risk, loss or probability weighting.
- **rank-dependent weights**. Decision weights that depend on where an outcome ranks among a gamble's outcomes, taken as differences of the weighting function at cumulative probabilities. Field: decision theory. Excludes weighting each outcome's probability on its own, as the original 1979 prospect theory did.
- **Prelec weighting function**. A two-parameter probability weighting function, exp of minus delta times minus log p to the power gamma, that sets the curvature and the crossing point of the curve separately. Field: decision theory. Excludes the one-parameter Tversky and Kahneman function used in CPT.
- **Shannon entropy**. Minus the sum of p log2 p over a gamble's outcomes, in bits: how uncertain the outcome is before it happens, 0 for a sure thing. Field: information theory. Excludes the sizes of the payoffs.
- **drive reduction**. For a setpoint S, the quantity S squared minus (S minus x) squared for an outcome x, whose expectation is linear in the mean and mean square of outcomes. Field: motivation psychology, our operationalisation. Excludes any measured need or state of the people choosing.
- **cross-validated R^2**. One minus the squared error of out-of-fold predictions over the variance of the observed rates, with 10 folds grouped by problem: the share of variance that predictions made without seeing those

rates account for. Field: statistics and machine learning. Excludes in-sample R^2, which rewards fitting noise.

- **quasi-binomial BIC**. The Bayesian information criterion with the log-likelihood divided by the dispersion: a score that trades fit against the number of parameters, lower being better. Field: statistics, model selection. Excludes an absolute test of a model, since it only compares models on the same data.
- **dispersion**. The factor phi by which observed rates scatter more than independent binomial choices would allow, taken from the COMB fit in each setting. Field: statistics, quasi-likelihood. Excludes a parameter of choice behaviour.
- **gradient-boosted trees**. A predictor built by adding many small decision trees, each fitted to what the previous ones missed; here a flexible learner's score on the same folds. Field: machine learning. Excludes a theory of choice, since it states no value function.
- **sampling ceiling**. One minus the mean sampling variance of the observed rates over their variance across problems: the share of variance not explained by finite samples of participants. Field: our usage, after reliability analysis. Excludes noise of other kinds, such as the structured decision noise reported for these data.
- **offset**. A part of the prediction fixed in advance and never refitted; here the fitted CPT log odds, so an added term can only explain what CPT leaves over. Field: regression. Excludes refitting CPT together with the added term.
- **iteratively reweighted least squares**. The standard routine for fitting a logistic model by repeated weighted least-squares steps. Field: statistics, generalised linear models. Excludes any effect on what is tested, since it is only the fitting method.
- **deviance**. Twice the log-likelihood shortfall of a model against a perfect fit; a deviance gain is the fall in deviance when a term is added. Field: statistics, generalised linear models. Excludes any penalty for the added parameter.
- **p value (from draws)**. The share of control draws scoring at or above the real score, with one added to the count and to the number of draws, so it is never zero. Field: statistics. Excludes a p value from a textbook formula that assumes a distribution.
- **Benjamini-Hochberg correction**. A rule that adjusts a set of p values into q values so that, among the tests called significant, the expected share of false calls stays at or below a chosen rate q, here 0.05. Field: statistics, multiple testing. Excludes the stricter Bonferroni rule, which bounds the chance of any false call.
- **95% interval**. The range that holds the middle 95 percent of the bootstrap or control draws for a number; we report it beside the number. Field: statistics. Excludes a statement that the true value lies inside with 95 percent probability.
- **bootstrap**. Refitting a model many times on problems resampled with replacement to read how much an estimate varies. Field: statistics. Excludes an interval from a textbook formula that assumes a distribution.
- **extra-binomial noise**. Scatter in a choice rate from problem to problem beyond what independent choices would give, simulated here with a Beta distribution. Field: statistics. Excludes a bias in the rate.
- **null distribution**. The spread of scores a control gives over many random draws; its 95th percentile is the bar a real score must clear. Field: statistics. Excludes a theoretical distribution assumed rather than drawn from the data.
- **false-success check**. A run of the full procedure on data simulated with no added effect, showing how large a gain the procedure finds when there is nothing to find. Field: our usage. Excludes a planted check, which asks whether a real effect is found.
- **findings ledger**. The append-only, dated list of every result we have recorded, numbered F1, F2 and on; corrections are added as new dated lines and old entries are never rewritten. Field: our records. Excludes a curated list of successes.

15 References

[1] Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. and Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. Science 372, 1209-1214. doi:10.1126/science.abe2629. Data: https://github.com/jcpeterson/choices13k

[2] Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297-323. doi:10.1007/BF00122574

[3] Brown, A. L., Imai, T., Vieider, F. M. and Camerer, C. F. (2024). Meta-analysis of empirical estimates of loss aversion. Journal of Economic Literature 62, 485-516. doi:10.1257/jel.20221698

[4] Prelec, D. (1998). The probability weighting function. Econometrica 66, 497-527. doi:10.2307/2998573

[5] Hertwig, R., Barron, G., Weber, E. U. and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. Psychological Science 15. doi:10.1111/j.0956-7976.2004.00715.x

[6] Thomas, T., Straub, D., Tatai, F., Shene, M., Tosik, T., Kersting, K. and Rothkopf, C. A. (2023). Modelling dataset bias in machine-learned theories of economic decision-making. Nature Human Behaviour. doi:10.1038/s41562-023-01784-6