Most people in private die and coin tasks do not claim the top number, and their reports cannot say why

A preregistered held-out comparison of six preference models on 27,039 subjects from 58 papers in the Abeler, Nosenzo and Raymond meta-study data

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F329 in the findings ledger. Signal (tier B).

Abstract

In a private die or coin task a person sees a random outcome alone and is paid by the number they report, so a lie pays and cannot be caught. Payoff maximisation says everyone should claim the top outcome. We used the open meta-study data of Abeler, Nosenzo and Raymond: 252 first-round treatments from 58 papers with 27,039 subjects, split by paper so that 19 papers with 9,960 subjects were held out. On held-out tasks with three or more outcomes, reports read as 45 percent honest, 26 percent maximal and 29 percent partial. Six preference models were fitted on the other papers and scored on the held-out ones. All six sit within 0.005 nats per report of each other, and no contrast against the fixed-cost model passes the false-discovery correction. The planted check failed: on data simulated from the reputation model, the reputation model lost to the fixed-cost model by 0.0003 nats per report. By the preregistered rule, no ranking among the models counts. Our subset's mean standardised report was 0.287 against the 0.234 published for the full data, outside the preregistered tolerance of 0.03. The departure from strict self-interest is large; the mechanism behind it is not identified by report distributions.

1 What this means for a reader

Most people given a free, safe chance to lie for money do not take all of it. Close to half report what they saw, about a quarter claim the top number and the rest shade upward partway. A model in which lying has a fixed cost, one in which people care about looking honest, one in which they follow what others do, and a plain payoff model with enough noise all reproduce those reports almost equally well. To learn which motive is at work, you need experiments designed to pull the motives apart. The distribution of reports alone is not enough.

2 Introduction

Fischbacher and Föllmi-Heusi [3] made the private reporting task a standard tool. Subjects rolled a die under a cup, reported the number, and were paid by it. No single lie could be detected, and the honest distribution is known, so lying shows in the reports as a whole. They found many fully honest subjects, a smaller group who lied to the maximum, and a large group of partial liars who lied but did not claim the top payoff, which they read as a wish to keep a favourable self-image.

Gächter and Schulz [4] ran the die task in 23 countries and related intrinsic honesty to a national index of the prevalence of rule violations built from corruption, tax evasion and political fraud. They found more honest reporting in subject pools from countries with fewer violations.

Abeler, Nosenzo and Raymond [1] combined data from 90 studies across economics, psychology and sociology and found that people lie surprisingly little. They derived predictions that separate the leading models, tested them with new experiments, and concluded that a preference for being honest together with a preference for

being seen as honest best explains the behaviour. They published their data [2].

This paper asks a narrower question of that public data. Without the designed treatments that Abeler and colleagues used to separate the models, can the plain report distributions of ordinary first-round tasks tell the models apart on papers the models have not seen? And what share of reports look honest, maximal or partial?

3 Preregistration and design

The design was written as a preregistration and hash-locked on 31 October 2026 at 22:02 UTC, before any model was fitted. Two things were seen first, and both are stated in the locked file. The publishers' site gives a mean standardised report of 0.234. Our parsing step printed pooled report distributions over all 252 treatments, held-out ones included, and printed a mean standardised report of 0.287. The forecasts of report shares are therefore not blind to the pooled shape.

The locked file names the data and its filters, the six models and two baselines, the score, the paper-level tests, the three-part reading, the planted check, a reproduction check, a control, the forecasts and the kill conditions. Its assumptions are listed in it: true outcomes are uniform; upward lies come from preference only and downward lies only from mistakes; one population spans countries and subject pools; only first rounds are used; and the norm model has no measured norm input. Nothing was added after the run.

Kill conditions, without discretion:

- If the planted check fails, no model comparison counts.
- If the best theory's gain over everyone-honest is not above the shuffle control's 95th percentile, the theories carry nothing beyond chance.

Reproduction check: the mean standardised report over the by-treatment set must lie within 0.03 of 0.234.

4 Data and provenance

The data are the merged meta-study file and the data-and-programs archive published by Abeler, Nosenzo and Raymond [2] on their resources page and on figshare (files 16211150 and 16211414), downloaded on 31 October 2026. The site states that all its data are released under the CC BY licence, with credit and a link back. The files were read as data; no program in the archive was run. The publishers' site describes the full set as 429 treatments from 90 papers with over 44,000 participants in 47 countries.

A parsing step kept first-round reports from tasks with uniform true outcomes and 2, 3, 6 or 10 possible outcomes. It removed treatments coded in the data as internal lying or as having a victim, treatments that asked subjects about the report distribution, and treatments that told subjects what others reported. That left 252 treatments from 58 papers with 27,039 subjects. Papers were split by a hash of the citation, with 30 percent held out: the held-out papers were 19 papers with 9,960 subjects, and fitting used 17,079 subjects.

As a check on the parsing, the baseline treatment of Fischbacher and Föllmi-Heusi in our file has counts of 38, 44, 64, 76, 171 and 219 from the lowest to the top payoff. Our results summary records the top report at 36 percent, where honest reporting gives 17.

5 Methods

Models

Let a task have K outcomes, ranked by payoff, with π_r the rank of report r scaled from 0 for the lowest to 1 for the highest. A person who saw outcome s and reports r receives the scaled payoff times a stakes factor

$$S = (G/5)^{1-\varepsilon} \tag{1}$$

where G is the money at stake in the treatment in purchasing-power dollars and ε is the stakes elasticity. Each person also carries a lying cost c, drawn from a lognormal distribution, $\ln c = \mu + \sigma z$ with z standard normal, evaluated on a grid of 25 quantiles. Each person picks the report with the highest utility

$$U = S\pi_r - ca_r - c(d_r)^\gamma + \rho c R_r \tag{2}$$

where a_r is 1 when r differs from s and 0 otherwise, d_r is the payoff distance $|\pi_r - \pi_s|$, and R_r is the chance that a person making report r is honest. Each model keeps some of these terms:

- Fixed cost: the first cost term only. A person either reports honestly or claims the top.
- Size cost: the second cost term only, so the cost grows with the size of the lie.
- Reputation: the fixed cost plus the value ρ on looking honest. Because R_r depends on everyone's reports, it is solved as a fixed point by 25 damped updates from honest reporting.
- Norm: the fixed cost scaled by $1 - \lambda L$, where L is the share of others who lie, iterated from no lying, which finds the low-lying equilibrium.
- Full: reputation plus the size cost.

A tremble τ mixes in uniform reports, so a model's prediction for report r is $q_r = (1 - \tau)m_r + \tau/K$, with m_r the share of the cost grid choosing r. The payoff model has no cost and no tremble. It is a logit choice on money alone with one scale parameter β:

$$q_r = \frac{e^{\beta G \pi_r}}{\sum_j e^{\beta G \pi_j}} \tag{3}$$

Two baselines frame the models: everyone honest, which predicts uniform reports, and the training-average baseline. Every model was fitted by maximum likelihood on the training papers.

Score

Models are scored by log loss over the N held-out subjects, with n_{tr} the count of report r in treatment t:

$$\ell = -\frac{1}{N} \sum_t \sum_r n_{tr} \ln q_{tr} \tag{4}$$

A model's gain is the everyone-honest loss minus its own. Each theory model is compared with the fixed-cost model by a paper-level sign permutation test, the p values adjusted together by the Benjamini-Hochberg correction. The reputation and full models' differences from fixed cost also get a 95% interval from a paper bootstrap.

Shares of honest, maximal and partial reports

For each task size of 3, 6 or 10, reports are pooled and given a three-part reading. With q_1 the share reporting the lowest outcome and q_K the share reporting the top,

$$h = Kq_1 \tag{5}$$

$$m = q_K - \frac{h}{K} \tag{6}$$

with h capped at 1, and the partial share is $1 - h - m$. The three task sizes are combined weighted by subjects. Two-outcome tasks are left out, since they have no partial lie.

Standardised report

For the reproduction check each report is mapped to $x_r = -1 + 2(r - 1)/(K - 1)$, and the standardised report is averaged over the T treatments without weights:

$$M = \frac{1}{T} \sum_{t} \sum_{r} \frac{n_{tr}}{n_t} x_r \tag{7}$$

6 Controls and planted checks

The planted check simulated report counts from the reputation model at known parameters on the same treatments, kept the same split by paper, fitted the fixed-cost, size-cost and reputation models to the simulated training papers and scored them on the simulated held-out papers. The reputation model had to beat both. The planted check failed: on data simulated from the reputation model, the reputation model lost to the fixed-cost model by 0.0003 nats per report. It beat the size-cost model by 0.0005. The fitted reputation parameters also missed the planted ones: the log of the value on looking honest, ρ, was planted at 0.2 and fitted at -0.56. The method cannot recover a known model at this sample size, so by the kill condition no ranking of the real models counts.

The shuffle control permuted report labels within each held-out treatment, and the best model's gain over everyone-honest was recomputed 300 times to give a null distribution. Its 95th percentile was -0.086 (mean -0.112). The real best gain of 0.072 lies far above it (p from draws 0.003). The theories carry information beyond chance, and this kill condition did not fire.

The reproduction check failed. Our subset's mean standardised report was 0.287 against the 0.234 published for the full data, outside the preregistered tolerance of 0.03. The two numbers come from different sets: ours uses first rounds only and drops several designs. Our filtered subset does not reproduce the publishers' headline figure, and the 0.287 was printed before the lock, so the outcome of this check was knowable when the plan was written.

7 Results

Table 1 and Figure 1 give the held-out losses. Every model beats everyone-honest by about 0.07 nats per report. All six sit within 0.005 nats per report of each other, and no contrast against the fixed-cost model passes the false-discovery correction. The reputation model's edge over fixed cost was 0.003 (95% interval -0.001 to 0.008, p 0.40), and the full model's 0.002 (-0.008 to 0.012, p 0.75). The training-average baseline, which has no theory in it, scored the same as the norm model to four places.

Table 1. Held-out log loss in nats per report on 9,960 subjects in 19 papers, and gain over everyone honest. The p value is the paper-level sign permutation test against the fixed-cost model, 5,000 flips; none passes the Benjamini-Hochberg correction.

Model	Loss	Gain	p against fixed cost
Reputation plus fixed cost	1.2484	0.0718	0.40
Payoff logit	1.2488	0.0714	0.83
Size-of-lie cost	1.2491	0.0711	0.83
Full	1.2493	0.0709	0.75
Fixed lying cost	1.2515	0.0687	reference
Norm	1.2519	0.0683	0.81
Training average	1.2519	0.0683	0.96
Everyone honest	1.3202	0	not tested

The payoff model was forecast to do worse than honest reporting. With its logit noise it did better, by 0.071 nats per report, and finished second. A model with no lying cost at all fits these reports almost as well as the cost models.

The fitted stakes elasticities were 0.89 to 0.96 across the five cost models, so the size of the payment hardly moved reports in these data.

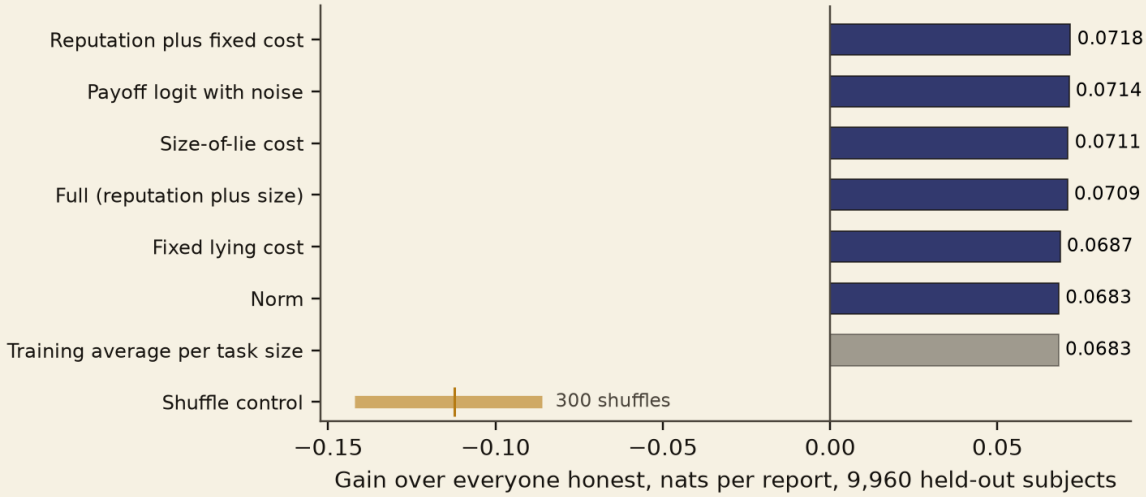


Figure 1. Held-out gain over everyone honest for each model (navy) and the training-average baseline (grey), against the shuffle control's 5th to 95th percentile range and mean (ochre). Drawn from the run's result file.

On held-out tasks with three or more outcomes, reports read as 45 percent honest, 26 percent maximal and 29 percent partial. The training papers gave 42, 24 and 34 percent. Table 2 and Figure 2 set those against the shares each fitted model predicts for the held-out tasks.

Table 2. Three-part reading of held-out tasks with 3, 6 or 10 outcomes, observed and as predicted by each model. The error is each model's mean absolute error over the three shares, by treatment, weighted by subjects.

Source	Honest	Maximal	Partial	Error
Observed, held-out papers	0.45	0.26	0.29	none
Observed, training papers	0.42	0.24	0.34	none
Norm	0.46	0.29	0.25	0.109
Reputation plus fixed cost	0.45	0.30	0.24	0.110
Fixed lying cost	0.49	0.29	0.22	0.116
Size-of-lie cost	0.36	0.32	0.32	0.124
Full	0.59	0.27	0.14	0.147
Payoff logit	0.61	0.16	0.23	0.160

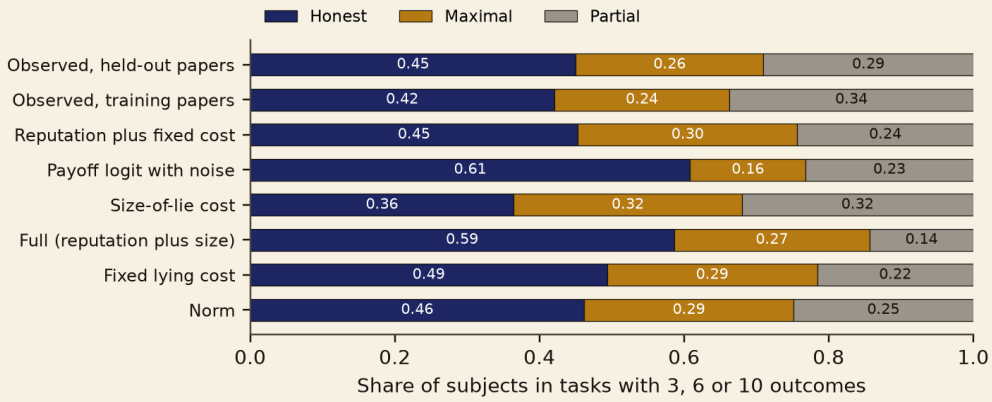


Figure 2. Honest, maximal and partial shares read from report distributions: observed on held-out and training papers, and predicted by each fitted model for the held-out tasks.

One held-out paper carries the full model's small edge. In Effron and colleagues' 2015 study (2,151 subjects), the full model beat fixed cost by 0.021 nats per report; without that paper, the full model's edge over fixed cost becomes -0.003. Across the 23 subject pools of Gächter and Schulz [4] in the data, the mean standardised report ranged from 0.18 to 0.58 (standard deviation 0.10).

Two sensitivity refits, scored on matching held-out subsets, are in Table 3. The fixed-cost model scored best of the three theories refitted in both. Among students the training average beat every theory; in the general population it lost to all three. The refit without hand-coded treatments was not run, because too few treatments remained.

Table 3. Held-out log loss after refitting on students only and on the general population only.

Subset	Subjects	Fixed cost	Reputation	Full	Training average	Everyone honest
Students	7,109	1.4285	1.4301	1.4338	1.4241	1.5188
General population	2,851	0.8067	0.8107	0.8067	0.8262	0.8249

Table 4 sets the forecasts written before the run against the outcomes.

Table 4. Forecasts fixed in the preregistration against what the run gave.

Quantity	Forecast	Outcome
Best theory model	full	reputation, within noise
Order	full, reputation, size, fixed, norm, payoff	reputation, payoff, size, full, fixed, norm
Payoff worse than everyone honest	yes	no, better by 0.071
Full over fixed cost	0.02	0.002
Training average over best model	0.005	-0.004
Stakes elasticity	0.9	0.89 to 0.96
Honest, maximal, partial	0.42, 0.25, 0.33	0.45, 0.26, 0.29
Reputation share error	0.10	0.110
Fixed-cost share error	0.20	0.116

8 Discussion

The departure from strict payoffs is the part that holds. Payoff maximisation predicts that every report is the top one; held-out reports show a top share of about a quarter beyond the honest share and a large honest share, consistent with what Fischbacher and Föllmi-Heusi [3] and Abeler and colleagues [1] report. The reading of those shares rests on the uniform-truth assumption and on there being no deliberate downward lies.

The mechanism is not identified here. Six models built on different motives, and a training average with no motive at all, predict held-out reports to within 0.005 nats per report. The planted check places part of the problem in the design: when the reputation model was true, the method still could not pick it out. Report distributions from ordinary tasks constrain the shape of reporting and leave the motive open.

This does not contradict Abeler and colleagues [1]. Their conclusion rests on designed treatments that change what a report signals about the reporter, which this study did not refit. Our result says that the distributions alone are not enough, which is the reason such treatments are needed.

The payoff logit's second place deserves care. A logit with a single scale parameter puts most weight on high reports and spreads the rest, and on these data that shape matches the reports about as well as a cost model. Its predicted shares are the furthest from the observed ones (Table 2), so a fit on log loss and a fit on shares can disagree.

9 Limitations

- The planted check failed, so the near-tie says nothing about which model is true, and the model ranking in Table 1 is not a result.
- The preregistered reproduction check failed: our subset gives 0.287 against the publishers' 0.234, and that value was printed before the lock.
- One held-out paper, Effron and colleagues (2015), carries the full model's edge; without it the sign flips.
- The shares are read from distributions, not observed lies. They assume uniform true outcomes and upward lies from preference only.
- The pooled report shapes were seen before the lock, so the share forecasts were not blind.
- All tasks are laboratory or online die and coin tasks with 2, 3, 6 or 10 outcomes, first rounds only.
- The designed treatments that separate the models were removed by our filters and not refitted.
- Scoring was done by our own code and was not blind.

10 Verified

Verified: on 9 November 2026 the subject and paper counts of the split, the eight held-out losses and gains, the permutation p values and their correction, the two bootstrap intervals, the shuffle control's draws and percentiles, the planted check's values and fitted parameters, the observed and predicted shares and the share errors, the stakes elasticities, the Effron sensitivity, the Gächter and Schulz range, the baseline counts, the sensitivity refits, the mean standardised report of 0.287 and all six verdicts were re-read from the run's result file. The 252 treatments, 58 papers and 27,039 subjects, the filters, the forecasts and the 0.234 comparison value were read from the preregistration; the publishers' full-set figures were read from our results summary of 31 October 2026. The results match the findings ledger and its dated correction of 9 November 2026. The preregistration's SHA-256 was recomputed on that date and matches the hash in the result file. The data source and licence were read from our download record. Both figures were drawn by a script that reads only the result file. The model forms in the methods were read from the study's code. The 36 percent top share was read from our results summary. The study was not re-run for this paper. The literature was found by searches run on 9 November 2026 and read at the level of abstracts, publisher pages and search summaries; the full texts were not read.

11 Corrections

- 9 November 2026: a dated correction in our findings ledger records that the preregistered reproduction check failed at 0.287 against 0.234 within 0.03. The original ledger entry did not state it. The claim stands narrowed, and this paper states the failed check in the abstract, the controls and the limitations.
- 9 November 2026: first published version.

12 Reproduction

- Data: download the merged meta-study file from the resources page of Abeler, Nosenzo and Raymond (figshare file 16211150), CC BY.
- Keep first-round reports from uniform tasks with 2, 3, 6 or 10 outcomes; drop internal-lying, victim, asked-distribution and information treatments; split papers 70 to 30 by a hash of the citation.
- Fit the six models of eqs. 2 and 3 by maximum likelihood on the training papers and score held-out loss with eq. 4.
- Read the three shares with eqs. 5 and 6, and the standardised report with eq. 7.
- Preregistration SHA-256: bff18a49df2fb811b8d8039a83d150a5a9d202a5e355ed8a4b79d5738c10f705.

Plan SHA-256: 76393962abab4576b991da0082fd66316b78ddcd9620080a11c4fc247ae73763. Run 20261031T220208Z.

- Cost: the analysis took about six minutes and the shuffle control under one second on one desktop computer, as timed in the result file.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run, and it passes only when every rule holds. Values are read from the run's result file. Because the planted check failed, the two model contrasts that pass are void as rankings, and neither passes the correction for multiple tests.

Table 5. Preregistered verdicts and the rules that decided each one. The reproduction check was preregistered as a rule and not coded as a verdict in the run; its value is the result file's mean standardised report.

Verdict	Deciding rule	Value	Holds
Planted check	reputation beats fixed cost and size cost on simulated reputation data	-0.0003; 0.0005	no
Beats shuffle	best gain above shuffle 95th percentile (-0.086)	0.072	yes
Payoff below honest	payoff gain over everyone honest below 0	0.071	no
Reputation over fixed cost	difference above 0	0.003	yes
Full over fixed cost	difference above 0	0.002	yes
Training average above best	training average's loss lower than the best model's	-0.004	no
Reproduction	mean standardised report within 0.03 of 0.234	0.287	no

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.

- **private reporting task (die-under-the-cup)**. An experiment in which a person observes a random outcome in private, such as a die roll or coin toss, reports it, and is paid by the report, so a lie pays and cannot be caught for any one person; lying shows only in the distribution of many reports. Field: experimental economics and psychology. Excludes tasks where the experimenter can see each person's true outcome.
- **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.
- **held-out papers**. Papers kept out of all fitting, chosen by a hash of the citation (30 percent of papers), on which every model is scored once; all treatments from one paper fall on the same side. Field: statistics, cross-validation. Excludes a split by treatment or subject, which would let one paper's design appear in both fitting and scoring.

- **stakes elasticity**. The parameter epsilon in a stakes factor (G/5) to the power 1 minus epsilon, where G is the money at stake; at epsilon 1 the size of the payment has no effect on reports, and at 0 the pull to lie grows in proportion to the payment. Field: behavioural economics, our parametrisation. Excludes the effect of payment on whether people take part.
- **fixed point (reputation model)**. A report distribution that, when used to compute how honest each report looks, leads people to make the same distribution of reports; found here by repeated damped updates starting from honest reporting. Field: game theory. Excludes a guarantee that the fixed point found is the only one.
- **tremble**. A parameter that mixes a share of uniformly random reports into a model's prediction, standing for mistakes and inattention; in these models it is the only source of downward lies. Field: game theory and behavioural economics. Excludes a preference, since the tremble does not depend on payoffs.
- **training-average baseline**. A forecast that gives each held-out treatment the average report distribution of training treatments with the same number of outcomes. Field: forecast verification, where it is called climatology. Excludes any theory of why people report as they do.
- **maximum likelihood**. Fitting a model's parameters by choosing the values that make the observed data most probable under the model. Field: statistics. Excludes choosing parameters by hand or by fit to the test data.
- **log loss (nats per report)**. Minus the average natural log of the probability a model gave to each report that was made; lower is better, and the difference between two models in nats per report is how much better one predicts each report. Field: statistics and forecast verification. Excludes an accuracy count, which ignores how much probability a model gave to what happened.
- **paper-level sign permutation test**. A test that sums each held-out paper's difference in log loss between two models, then flips the sign of each paper's difference at random 5,000 times to see how often a sum as large arises with no true difference. Field: statistics. Excludes a test that treats subjects within one paper as independent.
- **Benjamini-Hochberg correction**. A rule that adjusts a set of p 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.
- **paper bootstrap**. An interval made by recomputing a statistic 2,000 times on held-out papers drawn with replacement, so subjects from one paper stay together. Field: statistics. Excludes resampling single subjects as if they were independent.
- **three-part reading**. A split of a report distribution into honest reporters, read as the number of outcomes times the share reporting the lowest outcome, maximal liars, the excess at the top over the honest share, and partial liars, the rest. Field: experimental economics, after Fischbacher and Föllmi-Heusi. Excludes an observation of any person's lie; it assumes the true outcomes are uniform and that no one lies downward on purpose.
- **standardised report**. The mean report of a treatment after the outcomes are mapped evenly onto -1 for the lowest-paying report and +1 for the highest, so honest reporting gives 0 in expectation and everyone claiming the top gives 1; here averaged over treatments without weights. Field: experimental economics; the measure Abeler, Nosenzo and Raymond report, computed here by our own code. Excludes a share of liars, which the mean alone does not give.
- **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.
- **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.

- **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.
- **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.

15 References

[1] Abeler, J., Nosenzo, D. and Raymond, C. (2019). Preferences for truth-telling. Econometrica 87(4), 1115-1153. doi:10.3982/ECTA14673

[2] Abeler, J., Nosenzo, D. and Raymond, C. Preferences for truth-telling: data and programs. Resources page of www.preferencesfortruthtelling.com; figshare files 16211150 and 16211414. CC BY. Downloaded 31 October 2026.

[3] Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise: an experimental study on cheating. Journal of the European Economic Association 11(3), 525-547. doi:10.1111/jeea.12014

[4] Gächter, S. and Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. Nature 531, 496-499. doi:10.1038/nature17160