capstract:2610.00002v1 [econ.EM] 26 Oct 2026

Disclosure and outcome realization under limited discretion

Mina E. Harasawa[1,*] Youssef B. Amrane[2] Elena Sáez Lorquín[3]
Liv A. Norderby[1,3]
[1]Department of Econometrics, Rimsdal School of Economics, Bergen, Norway
[2]Institut d'Évaluation Sociale de Tazarine, Rabat, Morocco
[3]Centro de Métodos Aplicados, Instituto Valdefresno, Zaragoza, Spain

October 26, 2026

Abstract

In evaluating the effect of disclosure on the realization of a privately stated objective, it matters whether the recipient can determine the outcome and whether the outcome is observed independently of the originator's report. We study both conditions in a three-arm experiment with 1,080 originators. Each originator selects a target for two outcomes, one determined by a recipient and one fixed before disclosure. Routing the target to the decision maker instead of a witness increases objective realization for the discretionary outcome by 35.8 percentage points (95% confidence interval 28.9 to 42.7). The corresponding contrast for the fixed outcome is 2.2 points (−4.0 to 8.4). Reporting departs from objective realization even for the fixed outcome. Among originators whose fixed outcome matched their target, affirmative reports fall from 90.1% without disclosure to 74.4% with disclosure to a witness. The experiment distinguishes the effect of disclosure on implementation from its effect on the assessment of implementation.

Keywords: random assignment, disclosure, negative controls, outcome measurement, discretion

1 Introduction

Studies of private objectives often use the same person to supply the intention, characterize who learned of it, and determine whether it was achieved. This arrangement matters when revealing an intention changes the set of people in a position to act on it. It also matters when it changes the standard by which the originator calls a result successful. Even complete and accurate records of who spoke to whom do not separate those channels. A record of the intention taken before any disclosure, and a measure of the outcome supplied by a different source, make it possible to define the two contrasts.

We use the birthday wish as a private objective. The instruction to keep it to oneself is widespread in the communities represented in our sample, although participants disagree about whether there is a person one can tell, and when. Observational accounts record several rules simultaneously (Bennis and Kessou, 2018; Moe and Bakkeli, 2020). Some allow disclosure after the expiry of the wish. Others treat asking about the outcome as a request to disclose the original ob-

∗Corresponding author: m.harasawa@example.org
Preprint

ject. The survey distinction between a wish that failed and a wish whose maker can no longer report failure is consequently difficult to implement. The published follow-up materials in (Azrou and Montesa, 2021), for example, allow an observer to record that a candle was extinguished but supply no independent entry for the target.

Our study makes the target a single symbol with two opportunities for realization. One symbol is dispensed by a mechanical device whose output has been fixed in advance. A second symbol is selected by a person in an adjacent booth. A participant chooses one symbol as a target for both outcomes, then extinguishes a light on the table and commits the target in a private terminal. In one arm the record remains sealed. In another a witness receives it. In the third the decision maker receives it. The originator cannot learn which recipient, if any, was sent the target until both outcomes are locked. At that point we show the outcomes and identify the route, then record the originator's judgment separately for each outcome. The sequence permits disclosure to affect the discretionary implementation and allows disclosure to affect the judgment of either implementation.

We distinguish those objects using the potential-outcome formulation of (Rubin, 1974; Holland, 1986) and the experimental design principles in (Fisher, 1935; Gerber and Green, 2012). Distinguishing the person who has information from the person who can act on it is central in settings with peer effects (Manski, 1993; Aronow and Samii, 2017). Our experiment assigns the recipient and prevents communication between recipients. A restriction on the permissible causal paths then yields a negative-control outcome for realization and an ordinary outcome for reporting. Section 2 describes the routing rules, Section 3 states the estimands, and Section 4 gives estimates and the negative-control checks.

2 Design and measurement

2.1 Recruitment and roles

The experiment used 60 sessions, each with 18 groups of three adults. Bergen, Rabat and Zaragoza each held 20 sessions. We recruited an originator per group from adults whose birthdays fell within two weeks of the session. Other volunteers filled the roles of witness and chooser. Witness and chooser assignments were randomly allocated among the other volunteers after arrival, and those roles did not change during a session. One originator, one chooser and one witness occupied three separated bays. Each person participated in a single session. Payment was fixed, unaffected by any route, target choice or eventual agreement. The three site committees approved the study and all participants consented before role-specific instructions.

We described the goal as investigating private choices and the recording of outcomes. The originator practiced making a choice without sending a message. The witness learned that a message could arrive for inspection and would not call for a response. The chooser learned that they would select one of four symbols to return. We kept the set of symbols constant across roles, but permuted their spatial order independently for each display. Otherwise an originator might wish for the same corner that a chooser hits out of habit. The permutation was fixed before the target was chosen and could not expose it. The menu consisted of a square, an open circle, a diamond and a five-pointed star, each printed in black at the same nominal height. The four options were available even if a participant thought that one was customary.

Each session contributed six originators per route. We retained every originator with a target locked at randomization (Table 1).

Table 1: Assignment and preassignment target choices, by route. Receipt was verified from the recipient terminal. The sealed route had no recipient. Each route contributes 360 recorded pairs of outcomes.

Route	Assigned	Target square	Open circle	Diamond	Star	Receipt verified
Sealed	360	91	78	80	111	n/a
Witness	360	85	82	78	115	354
Chooser	360	92	75	81	112	351
Total	1,080	268	235	239	338	705

2.2 Initial record and transmission

The originator selected a target on a private terminal. With the target still on screen, they turned off the light and pressed the commit key. Before this key was pressed, the terminal treated the displayed choice as editable. Only the committed target entered the analysis. It was possible to think of another object privately. The object we could follow in the data was the one for which a person requested recognition on the terminal. A pilot compared recording the target before and after extinguishing the light (Harasawa and Amrane, 2024). Recording it afterwards allowed substitutions after the initial ritual. We used the earlier position in the sequence and added the commit key to distinguish candidate edits from the final entry.

The terminal allocated each linked group to one of the 18 positions on a previously randomized list for the session. The assignment program revealed the route only to the message broker once all targets in that session were locked. Removing a participant before that point left the entire list unused until a replacement could complete the baseline procedure. No replacements occurred after the list was unsealed. The broker either routed a message containing the symbol and the statement “This is the stated target for this trial” or held the message in an inaccessible record. Both recipients saw the same header followed by either the symbol and statement or a neutral notice saying the record would be available after the study. The witness could acknowledge a message but had no input bound to the symbol selection. The chooser was free to pick any symbol, including one different from that stated in the message. A four-second minimum display time was common to the two recipient terminals.

The identity of the birthday participant was withheld from both recipients. In three sessions, attendants prompted choosers who reached the time limit. The screen and message state stayed fixed.

2.3 Outcome log and report

The mechanical device contained four-symbol output slips of equal probability. A separate process assembled one independent draw per group before its route list was unsealed. A container with that slip was locked until the chooser submitted a symbol. The machine then dispensed the slip and posted its symbol to the originator’s screen. The study logged the fixed symbol when it loaded the container and again at presentation. The software equated the two codes before opening the report screen. The chooser had no control over dispensing or its content. The complete script and time order are described in Appendix A.

The originator then saw both realized symbols, with their sources labeled. The feedback panel identified who received the record, using the allocated route even if the other terminal had failed to confirm receipt. We chose this wording in advance because confirmation can occur after the

outcome selector starts. The report question named the two physical sources separately and asked "For this outcome, did your wish come true?" It accepted yes or no. There was no scale, auxiliary question or skip between the two report items. Order of the two items was randomized independently of route. For each source the display showed the symbol produced and kept the target off the feedback panel. The retained target allowed us to score realization without asking the participant to remember it later. The screen did not equate the intended criterion for the participant with ours. Lovann and Pellin (2022) discusses this distinction between outcomes recorded by the researcher and correspondence judged by the subject, including settings with identical input events.

Everyone responded to both questions. As a late comprehension check we asked the originator to identify the assigned route and recall the target symbol. Successful recall of target was 96.3% overall (1,040 of 1,080), and 1,006 correctly identified the route. Incorrect answers did not change the score or cause exclusion. Recall and reporting could both reflect the assignment, so selecting records with correct answers would change the populations being compared. In Section 4 we report bounds where recall errors are given their most adverse possible interpretation.

3 Estimands and inference

3.1 Realization and reporting

Let Z_i be route S for sealed, W for witness or C for chooser. The sources are H for chooser and M for the mechanical device. Write A_i for the fixed target, $O_{ik}(z)$ for the selected symbol for source k and $Y_{ik}(z)$ for the resulting realization indicator. $R_{ik}(z)$ is the originator's report. The realization contrast for route a against b is

$$Y_{ik}(z) = \mathbf{1}\{A_i = O_{ik}(z)\}, \qquad \tau_k(a,b) = \frac{1}{N}\sum_{i=1}^{N}\big(Y_{ik}(a) - Y_{ik}(b)\big). \tag{1}$$

These potential outcomes are defined for linked groups. The design assumes no interference between groups. If outcomes for another group depended on a recipient's assignment, the marginal rate would instead average over assignments to those groups (Aronow and Samii, 2017). Separation of bays, prohibition of shared recipient terminals and a single exposure per participant limit such dependence. We cannot exclude anticipation based on experience outside the lab, which is common to the arms in expectation and may modify the effect of route. In practice the potential outcomes include the recipient's response to the message as delivered by this protocol, even when it is not read. Imbens and Angrist (1994) explains the additional conditions needed to recover an effect of receipt among compliers. In our setting assignment is also displayed to the originator at reporting. For R an exclusion restriction based solely on receipt is particularly implausible.

The primary contrast is $\tau_H(C,W)$, which changes whether the person shown the target can select the human outcome. The route contrasts C against S and W against S are secondary. $\tau_M(C,W)$ concerns the same change of route for a symbol determined before assignment and scored against a target also fixed before assignment. If the container and log behave as specified, $Y_{iM}(C) = Y_{iM}(W) = Y_{iM}(S)$ for every group. This is a sharp null for mechanical realization under the maintained design, irrespective of whether targets are uniformly distributed. A randomization test there checks the alignment of recorded assignments and outcomes. A wide confidence

interval would remain compatible with material implementation deviations. The design gives the mechanical null more force than an imprecise estimate alone.

The difference in the realization contrasts uses $Y_{iH} - Y_{iM}$ as the outcome. It does not require assigning the two outcomes independently. We include it to show what is estimated when an investigator uses measured realization of the mechanical source to remove a shared baseline chance of success. That subtraction is unnecessary for identification of $\tau_H(C, W)$, and it may worsen precision. The adjustment does not in general recover the mechanism if the comparator itself responds to disclosure.

For reporting, R_{ik} is measured after feedback. The contrast of $R_{iM} - Y_{iM}$ captures a movement in signed disagreement, with positive values indicating excess affirmative reports. Within the stratum $Y_{iM} = 1$ the indicator $1 - R_{iM}$ measures a missed affirmative report. Since the mechanical indicator is known before route, conditional analysis there does not condition on a postassignment collider. The analogous exercise with $Y_{iH} = 1$ has no such interpretation: the chooser can change who belongs in that stratum. The distinction matters even for identical questions asked of the same people.

3.2 Estimation

Assignment probabilities equal one third for each route. We use equally weighted session differences, which, with six observations per arm in each session, are also the simple differences of the pooled means. They coincide with the inverse-probability estimator under this design (Horvitz and Thompson, 1952). For a given comparison let $\hat{d}_j$ be the difference in session j. We use variation among the 60 session differences as a conservative variance estimator for the realized population when sessions differ in their true effects. Appendix B gives the argument and the interval formula. We checked the sharp null for the primary and mechanical contrasts against 19,999 permitted assignment draws using the statistic in Eq. (2). Rearrangement kept witness, chooser and sealed counts at six within each session. Baseline covariate adjustment used the fully interacted regression approach in (Lin, 2013). Recruitment site and the four target indicators were known before routing, but only target indicators vary within the session blocks.

All confidence intervals in Table 2 and Figure 2 are pointwise. For the three secondary tests on realization reported with adjusted values in the protocol supplement, we used Holm's procedure (Holm, 1979). The primary test remains separate. Reporting outcomes were prespecified as a descriptive family and interpreted using their intervals. We give no cut-off at which the convention of secrecy counts as invalidated.

4 Results

4.1 Routing and objective realization

There were 1,080 committed targets and 2,160 observed realization indicators. The route confirmations in Table 1 show 705 acknowledged messages out of 720 transmitted. Thirteen missing confirmations came from timeouts and two from button failures recorded by the attendant. No sealed entries had any transmission record. Over 1,080 groups the presentation code for the fixed outcome always agreed with the code at loading. All outcomes and analysis rows were present.

The assigned chooser arm produced 225 objective successes at the discretionary source out of 360 (62.5%). The witness arm produced 96 (26.7%). The difference of 35.8 percentage points has a 95% interval of 28.9 to 42.7, and the blocked randomization test gives $p < 0.001$ against the sharp null. The sealed arm produced 90 successes (25.0%). That arm supplies the comparison in

which the participant's target was never made available outside the terminal. The witness-to-sealed difference at the chooser was 1.7 percentage points (−4.8 to 8.1). The contrast measures the effect of assigning a route. Individual chooser motives are unobserved.

The mechanical outputs realize 91, 86 and 94 targets in the sealed, witness and chooser arms. The chooser-to-witness contrast is 2.2 points (−4.0 to 8.4). Table 2 gives both realizations and reports, letting readers compare the two sources at each route. Figure 1 displays their levels on a common probability scale. The horizontal mark is one-quarter, the success rate for a uniformly selected output regardless of the originator's distribution of wishes. Without disclosure, chooser preferences may give a different rate.

Table 2: Objective target matches and self-reported realization by source and allocated route. Each cell in the first three rows has 360 observations. Contrast entries are percentage points with pointwise 95% intervals.

Route or contrast	Chooser matches	Machine matches	Chooser reports	Machine reports
Sealed	90 (25.0%)	91 (25.3%)	87 (24.2%)	86 (23.9%)
Witness	96 (26.7%)	86 (23.9%)	83 (23.1%)	66 (18.3%)
Chooser	225 (62.5%)	94 (26.1%)	207 (57.5%)	70 (19.4%)
Chooser less witness	35.8 (28.9 to 42.7)	2.2 (−4.0 to 8.4)	34.4 (27.7 to 41.2)	1.1 (−4.5 to 6.7)
Witness less sealed	1.7 (−4.8 to 8.1)	−1.4 (−7.6 to 4.8)	−1.1 (−7.5 to 5.3)	−5.6 (−11.4 to 0.3)

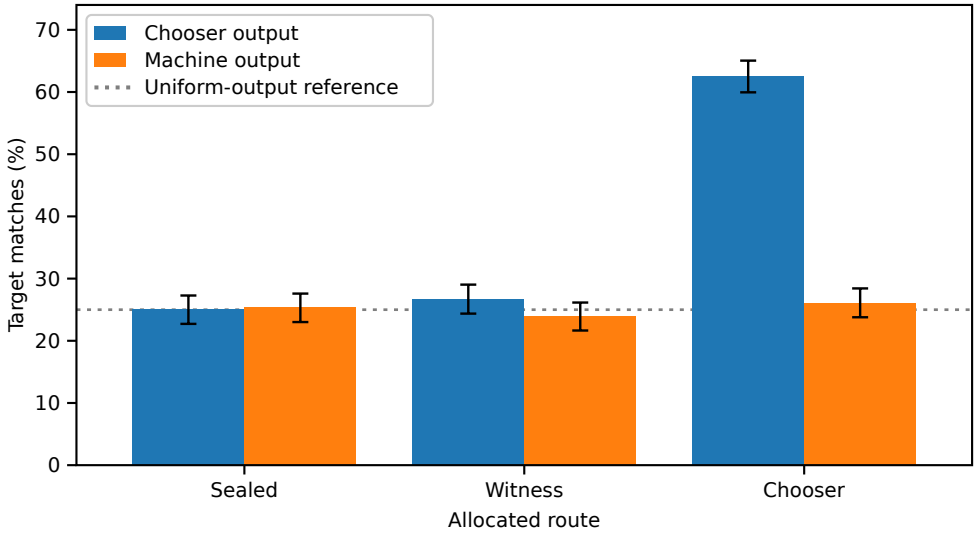


Figure 1: Objective target matches by route. Bars show binomial standard errors for the displayed unconditional rates, for scale. Contrast inference uses assignment blocks.

Subtracting the corresponding mechanical contrast from the primary contrast leaves 33.6 points (24.5 to 42.7). The interval is wider on allowing variation from the negative control, as visible in Figure 2. The baseline-adjusted primary estimate is 35.7 points (28.8 to 42.6). Covariate adjustment had little effect on the interval in this design, where treatment counts are already balanced within session.

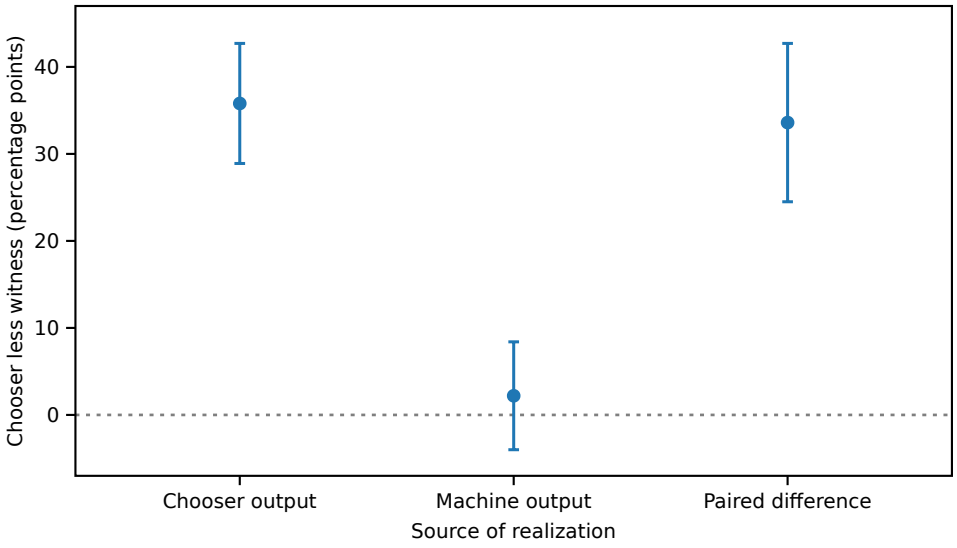


Figure 2: Assignment contrasts for objective realization. Bars give pointwise 95% intervals accounting for blocking. The paired difference uses one outcome per group equal to human minus mechanical realization.

4.2 Reporting when the target was met

The machine outcome makes the meaning of a negative report especially clear. Of the 91 objectively matched outcomes under sealing, 82 were called successful, an affirmative rate of 90.1%. Of the 86 matched outcomes under the witness route, 64 were called successful, a rate of 74.4%. Of the 94 under the chooser route, 68 were called successful, a rate of 72.3%. Reports without a target match are rare under all three routes. Table 3 cross-tabulates report status and mechanical realization. Figure 3 shows the implied rates with a denominator defined by the retained target and the fixed outcome. A published count that treated only reports as outcome data would miss 57 of the 271 mechanical successes, while adding eight successes that do not match the record.

Table 3: Report status for mechanical outcomes by route and objective realization. Objective success and failure refer to the preassigned device output and the locked target.

Route	Objective successes	Yes if success	Objective failures	Yes if failure
Sealed	91	82	269	4
Witness	86	64	274	2
Chooser	94	68	266	2
Total	271	214	809	8

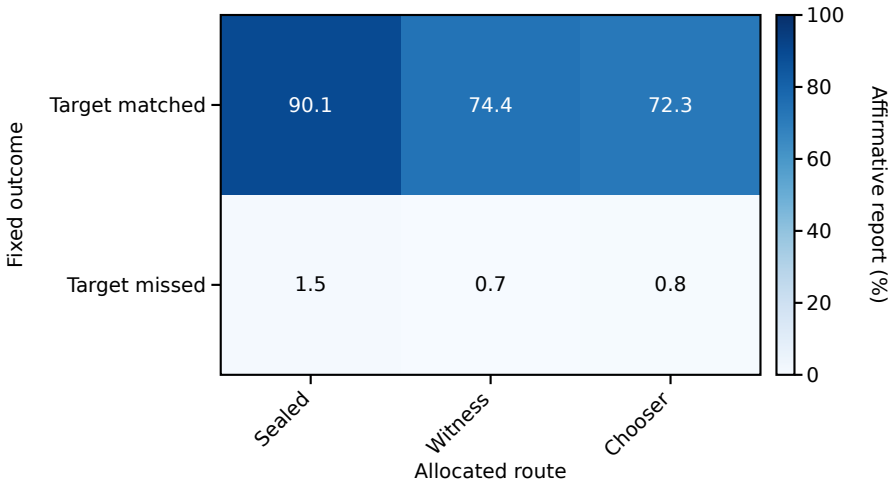


Figure 3: Affirmative reporting rates for mechanical outcomes conditional on objective match. The denominators are all participants at the indicated route and match status.

The difference in signed disagreement ($R_{iM} - Y_{iM}$) between witness and sealed arms is -4.2 points. The confidence interval is -7.2 to -1.1, using one difference per originator and the same assignment blocks. Comparing yes rates only among mechanically successful participants gives a decline of 15.7 points. This is the difference of two ratios with random denominators even though success itself was fixed. Both estimates permit subjects to interpret disclosure as failure of the procedure after seeing the desired result. They also permit reduced effort when entering the answer, including failures to retrieve the intended symbol from memory. The question asks whether the wish came true. It does not ask whether the wish had a hand in bringing about the outcome.

Recoding any of the 40 answers from originators who later misrecalled the target could move the signed contrast by at most $40/360$, or 11.1 points. This conservative bound includes zero even if all other reports are taken at face value. We keep the recalled and locked symbols separate because replacing the locked target would choose the criterion after the event (Rekdal and Ferrón, 2019).

5 Interpretation

A chooser can make a request materialize when given the target. A witness can receive the target and leave realization essentially where it was. The experiment locates those treatments separately. In ordinary practice the observer, potential helper and subsequent assessor may be the same person. Which role someone occupies is then partly determined by the disclosure. An observational estimate comparing wishes a person chose to share with wishes kept private can reflect both differences in targets and differences in later help. Conditioning on a record of that help could remove part of the effect one is trying to measure. Rosenbaum (2002) treats sensitivity to hidden bias in observational contrasts. No bound on such bias supplies outcomes that the study did not collect.

There is a smaller difference between the decision maker's opportunity to implement an objective and the opportunity to be credited with its implementation. The decision maker cannot affect the mechanical symbol, but the rate of affirmative reporting falls even when the desired symbol comes out of the device. The effect on reporting could reflect a distinction participants make

between receiving what they wished for and a wish coming true. Kallel and Veldén (2023) found that some respondents explicitly use the second expression only when they judge that the intended procedure remained valid. We elicited no explanation until after the binary judgments, and the survey did not ask whether a participant recognized a direct causal mechanism. The data therefore do not measure adherence to any particular explanation of success. The direct record tells us what a study using self-reports alone would count.

The symbol pair is a narrow class of objectives. It is costless for the chooser to select any outcome and there is no penalty to granting a wish. People contemplating a choice with real costs might react less to the request, or react differently to a stranger who can identify them. There is no requirement that a wish in the field be resolvable, possess one clear deadline or even still be wanted when the outcome occurs. Our design needed to hold each target fixed across two mechanisms to score both. That requirement precludes wish revision in response to an encouraging sign. The estimate addresses a stated target, given one chance to occur in each source, with each report made after both opportunities have ended. Neither secrecy before expiry nor retrospective disclosure after expiry is the same treatment as the witnessed arm we study.

The study discloses wishes without asking originators to tell someone. A voluntary disclosure arm could change both who shares a wish and what the recipient takes the message to mean.

Data availability

The assignment and outcome panel contains target and output codes but no text about events outside the session. The analysis packet, including the display sequence, the assignment masks and the rule for the fixed-output log, is held at the Rimsdal School of Economics. The public export suppresses birthday dates and the dates of individual sessions.

A Timing and record reconciliation

Each group had a session identifier and an ordinal within session. A record was written when the originator committed the target, when the broker released each terminal payload, when the chooser confirmed a selection, and when the scoring screen displayed both symbols. Preparation of the fixed output used a separate random stream and preceded broker release. Each slip container had a scanned code supplied to an attendant from a list ordered by group ordinal. An attendant could replace an unreadable container code before the trial, in which case the trial had to return to the unassigned state. On the data retained here, all loaded codes matched printed slips. The reconciled event stream preserves absence of acknowledgement as a field, distinct from absence of transmission. The two button failures were entered by site staff from the device log. In no case was an unlocked target visible to the chooser.

The randomization test kept the six positions per route and recomputed the estimator after moving the route labels within each session. It held targets and both outcomes fixed, as appropriate to the sharp null for the source under test. With b draws where the absolute statistic equals or exceeds its observed value, we report $(1 + b)/20{,}000$.

B Variance over sessions

Let $J = 60$, $n_j = 18$ and $n_{ja} = n_{jb} = 6$. Under the design the difference in the two sample means in block j estimates the mean finite-population difference d_j in that block. The assignment lists

are independent. The unadjusted estimator and the conservative variance estimate are

$$\widehat{\tau} = \frac{1}{J}\sum_{j=1}^{J} \hat{d}_j, \qquad \widehat{V} = \frac{1}{J(J-1)}\sum_{j=1}^{J} (\hat{d}_j - \widehat{\tau})^2. \tag{2}$$

Because the lists are assigned independently, $\mathrm{Var}(\widehat{\tau}) = J^{-2}\sum_j \mathrm{Var}(\hat{d}_j)$. Expanding the sum of squares in Eq. (2) gives

$$\mathrm{E}[\widehat{V}] = \mathrm{Var}(\widehat{\tau}) + \frac{1}{J(J-1)}\sum_{j=1}^{J} (d_j - \tau)^2. \tag{3}$$

The second term is nonnegative and equals zero when the mean effect is constant across sessions. It is a contribution from effect heterogeneity, distinct from the nonidentified covariance of two potential outcomes within a group in the ordinary complete-randomization formula (Neyman, 1990). We report approximate intervals with the t critical value for 59 degrees of freedom. Their coverage with 60 fixed blocks depends on concentration of the variance over blocks. The sharp-null tests need neither this approximation nor a constant-effects model. If two observed indicators are combined, for example into the within-group difference plotted in Figure 2, they are combined before forming the block differences. Variances of two separate estimates cannot be added without their covariance.

The within-realization reporting ratios use the pooled number of successful outcomes by route as denominators. We show their empirical values in Figure 3 without bars. Those ratios would generally require variance terms for the number of successes in each assigned arm, even under the mechanically invariant realization assumption. The simple session difference estimator is used for the signed difference in disagreement rates over all groups. This keeps the definitions of the inference targets separate.

Acknowledgments

We thank the site attendants for maintaining separate records of message receipt and outcome presentation. The Skarnes Fund for Experimental Economics covered participant payments.

References

Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics*, 11:1912–1947, 2017.

Kamil B. Azrou and Yolanda Montesa. Recording the end of an individual petition. Technical Report 21-07, Institut d'Évaluation Sociale de Tazarine, 2021.

Meryem A. Bennis and Rania B. Kessou. Domestic observance and the custody of private intentions. Technical Report 18-04, Centre d'Archives Sociales de Mers Louata, 2018.

Ronald A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.

Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton, 2012.

Mina E. Harasawa and Youssef B. Amrane. Response timing and target stability at commitment. Technical Report 24-16, Rimsdal School of Economics, 2024.

Paul W. Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81:945–960, 1986.

Sture Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

D. G. Horvitz and D. J. Thompson. A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

Guido W. Imbens and Joshua D. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62:467–475, 1994.

Amira Y. Kallel and Elise A. Veldén. Conditions on success in retrospective household interviews. Technical Report 23-09, Centre d'Archives Sociales de Mers Louata, 2023.

Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7:295–318, 2013.

Tiril E. Lovann and Jorge M. Pellin. Outcome correspondence with an independently recorded event. Technical Report 22-05, Havset Institute for Social Measurement, 2022.

Charles F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60:531–542, 1993.

Lise Randine Moe and Jarl A. Bakkeli. Secrecy and expiry rules in household celebration records. Technical Report 20-11, Havset Institute for Social Measurement, 2020.

Jerzy Neyman. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5:465–472, 1990.

Simen Lars Rekdal and Eva I. Ferrón. Prior records and post-event revisions of individual aims. Technical Report 19-02, Instituto Valdefresno, 2019.

Paul R. Rosenbaum. *Observational Studies*. Springer, 2002.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.