



KEN RESEARCH · METHODOLOGY

# Which persuasion science survived the replication crisis?

The data, the methods, every number and every limitation behind the study.

**Cristian Frunze**  
Founder of Ken

## ABOUT THE DATA

Published research only: 132 sources from 836 logged queries across three literature databases (Crossref, Europe PMC and Semantic Scholar), plus known-item lookups and citation chasing. No Ken internal data was used.

No personal data. Every number comes from a published source, listed in the methodology PDF.

Figures 1 to 3 are the study page's charts, drawn from the same numbers. The methodology after them keeps the full tables.

## SUMMARY

# Which persuasion science survived the replication crisis?

Of 12 famous persuasion effects sales playbooks cite, 4 held up in large tests. Those tests found a median 25% of the original effect.

### 4 of 12

effects replicated: anchoring, reciprocity, choice overload and name personalization. 6 were mixed, 1 failed and 1 was never tested in a large study.

### 25%

median size of the large-test effect against the classic study, across the 7 effects where both could be compared.

### 6 of 7

comparable effects were smaller in the large test. Anchoring was the only one that came out bigger.

### d = 0.00

for decoy pricing in its one large real-world test, more than 140,000 flight-booking sessions. The only effect graded Failed.

## Data and method

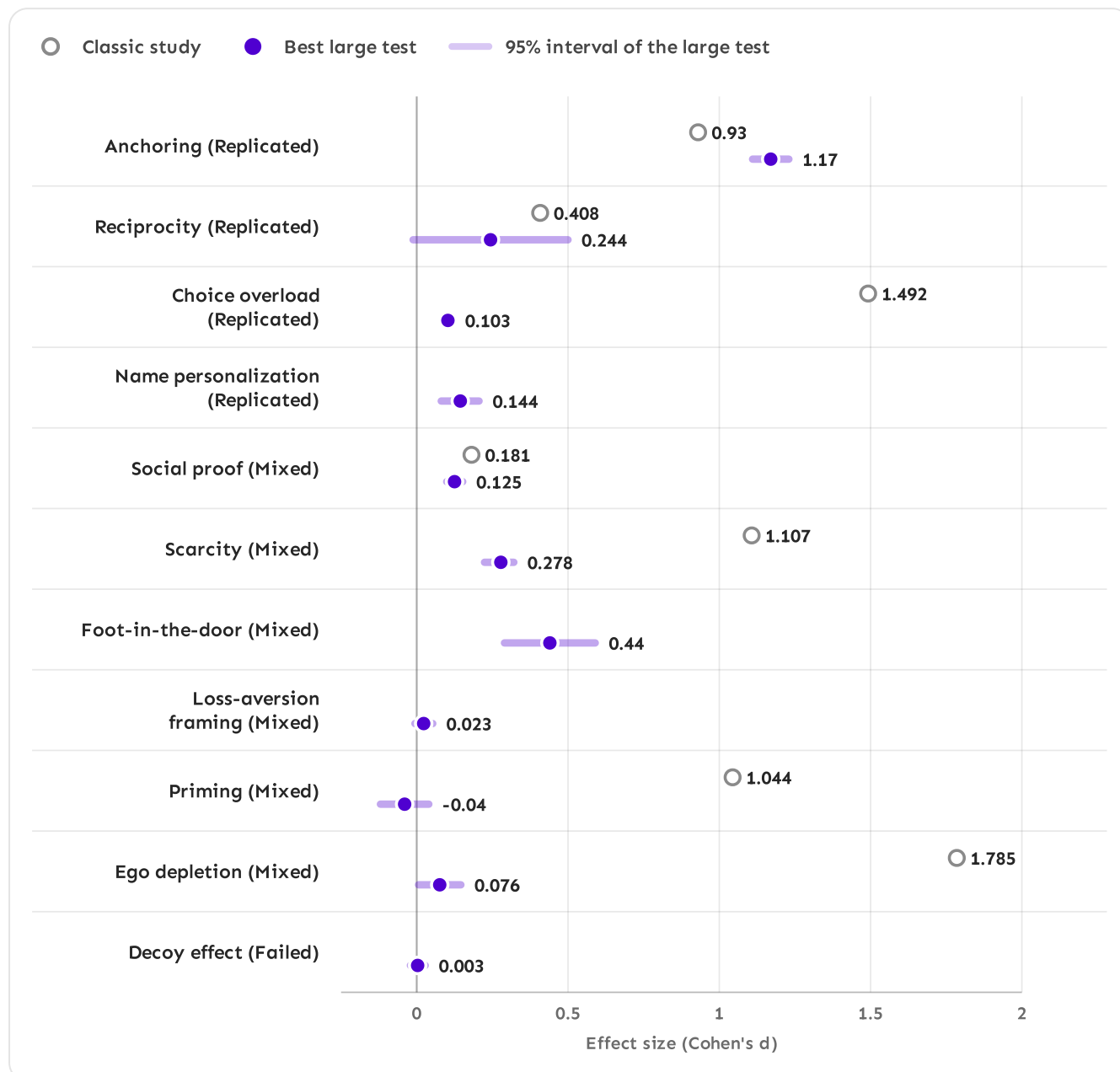
|                   |                                                                                                                                                                                                                                                                              |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Data</b>       | Published research only: 132 sources from 836 logged queries across three literature databases (Crossref, Europe PMC and Semantic Scholar), plus known-item lookups and citation chasing. No Ken internal data was used.                                                     |
| <b>Sample</b>     | 12 persuasion effects cited in sales and marketing playbooks, graded on 132 included sources.                                                                                                                                                                                |
| <b>Date range</b> | Databases searched on 15 October 2026. Classic studies from 1975 to 2008.                                                                                                                                                                                                    |
| <b>Method</b>     | Pre-registered evidence review: predicted grades, grading rules and effect-size conversions were committed before any search. A second reviewer checked every number against its source, and a fresh analysis recomputed every grade.                                        |
| <b>Excluded</b>   | <ul style="list-style-type: none"><li>• Narrative reviews, single small studies, books, press and vendor reports</li><li>• Preprints, except multi-site replications</li><li>• Studies below the strong-test thresholds, which give context but do not set a grade</li></ul> |
| <b>Privacy</b>    | No personal data. Every number comes from a published source, listed in the methodology PDF.                                                                                                                                                                                 |

## FIGURES

# The study's charts

FIGURE 1

## Classic studies against the best large tests

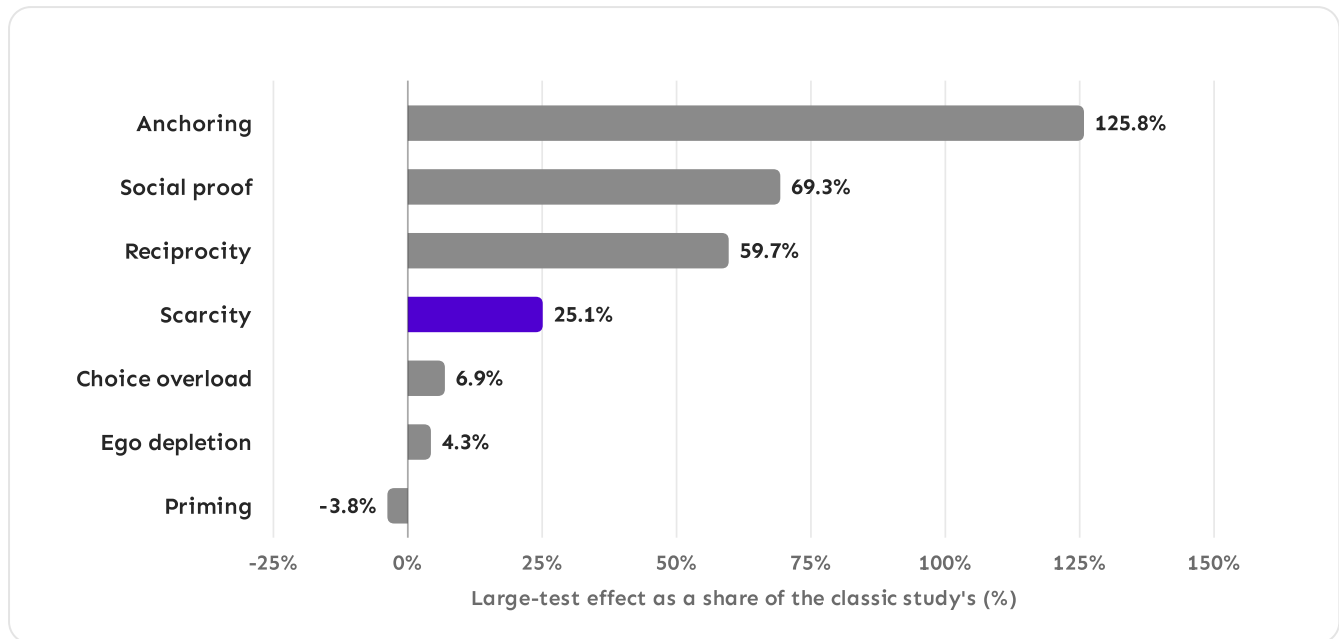


Each row is one effect, with its grade. The open circle is the classic study's effect size; the filled dot and line are the best large-sample estimate and its 95% interval. Cohen's d: 0.2 is small, 0.5 medium, 0.8 large. Four effects have no classic study to compare; the "because" effect has no large test.

Source: Ken Research review of 132 published studies, analysed by Cristian Frunze.

FIGURE 2

## Large tests found a median 25% of the original effect

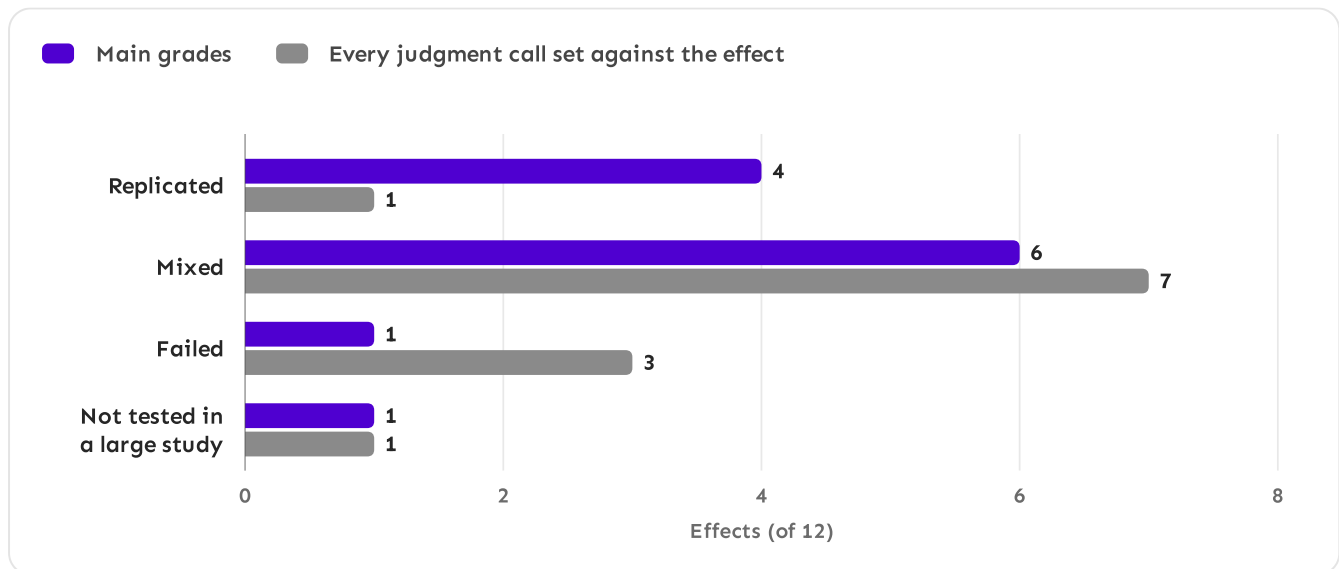


The best large-test effect as a share of the classic study's, for the 7 effects with both. 100% would mean the large test found the full original effect; below zero, the large test pointed the other way. Scarcity, in purple, is the median.

Source: Ken Research review of 132 published studies, analysed by Cristian Frunze.

FIGURE 3

## The grades, and how far they move on the judgment calls



Number of the 12 effects in each grade. The gray bars set every flagged judgment call against the effect: the most skeptical reading of the same evidence.

Source: Ken Research review of 132 published studies, analysed by Cristian Frunze.

# The data, the methods and every number

Literature review; no Ken internal data was used. All numbers below are read directly from the study's outputs.

## 1. Question and design

Twelve persuasion effects that sales and marketing playbooks cite were graded against the strongest tests published since the replication crisis: preregistered multi-site replications, large or preregistered field experiments, meta-analyses and large direct replications. Each effect gets one of four grades (Replicated, Mixed, Failed, Not tested at scale), a best large-sample effect size in Cohen's  $d$  with its 95% CI, a shrinkage ratio against the classic study playbooks cite, and a B2B transfer rating. Each effect is graded on its tactic version, the form a sales team would use; other versions are context.

---

## 2. Pre-registration

The hypothesis, predicted grades, grading rule, inclusion rules, effect-size conversions and falsification criteria were pre-registered in the study plan and protocol before any search ran. Predicted totals were 5 Replicated, 4 Mixed, 2 Failed and 1 Not tested at scale, with a median shrinkage ratio of 0.50 or less. The hypothesis would be wrong if 8 or more effects replicated, if priming or ego depletion replicated, or if the median shrinkage ratio exceeded 0.75. Operational rules added after the search are logged with dates and reasons in the study plan and listed in section 13 below.

---

## 3. Sources and search

- Databases: Crossref, Europe PMC (which covers PubMed and PMC) and Semantic Scholar, queried through their public APIs on 2026-10-15. Each of the 12 effects' term sets was combined with design terms (meta-analysis, registered replication, many labs, multisite, replication, field experiment); the full strings are set out in section 3 of the protocol. Crossref returned the top 20 hits per query and Europe PMC the top 50.
- The Semantic Scholar pass stopped after 44 of 125 queries because of rate limits. Totals: 3,428 hit rows before deduplication.
- Extraction and review agents then ran known-item lookups of the named replication programs (Many Labs 1-5, Registered Replication Reports, other multi-site projects), targeted searches, and backward and forward citation chasing for the strongest sources per effect. Query counts by purpose: main database search 262, targeted search 217, known-item lookup 268, backward citation chasing 47, forward citation chasing 42; 836 queries in all.
- Not available: OpenAlex (no API key configured), PsycINFO, Web of Science and Scopus.

---

## 4. Screening

Records were deduplicated by DOI, then by normalized title. A keyword pre-screen excluded records whose abstract contained no eligible design term; records without an abstract were always forwarded. The pre-screen first missed titles written with a Unicode hyphen (for example 'meta-analysis'); this was fixed and the affected records were screened by hand. Title and abstract screening was done by language-model agents with a fixed prompt and structured output, split by effect group. Every possible include then had a full-text or eligibility check, recorded with its reason.

| Stage                                                         | Count |
|---------------------------------------------------------------|-------|
| hit rows retrieved (all sources)                              | 3,428 |
| unique records after deduplication                            | 2,637 |
| excluded at pre-screen (no design term)                       | 604   |
| forwarded to title/abstract screening                         | 2,033 |
| unique records screened on title and abstract (all groups)    | 2,034 |
| excluded at title/abstract                                    | 1,895 |
| title/abstract candidates                                     | 139   |
| full-text / eligibility decisions recorded (unique DOIs)      | 386   |
| excluded at full text                                         | 254   |
| included sources (unique DOIs)                                | 132   |
| included, found via citation chasing                          | 12    |
| included, found via known-item lookup                         | 50    |
| included, found via screening                                 | 50    |
| included, found via targeted search                           | 20    |
| evidence rows extracted                                       | 189   |
| rows verified or corrected by the reviewer                    | 180   |
| rows dropped as unverifiable                                  | 9     |
| rows marked include_in_grade after review (any role or level) | 140   |
| unique sources behind those rows                              | 98    |
| queries logged (all passes)                                   | 836   |

## 5. Inclusion rules and evidence levels

A source was included when it studied humans, tested one of the 12 effects, reported an effect size with a 95% CI (or statistics to compute one), was published in English (preprints only for multi-site replications, flagged), and had one of these designs:

- **Level 1:** a preregistered multi-site replication with total N of at least 1,000, or a field experiment of the tactic with at least 10,000 units or a preregistration.
- **Level 2:** a meta-analysis with at least 10 effect sizes. A bias-adjusted estimate is preferred when reported, in the order: the authors' preferred adjustment, PET-PEESE, selection models, p-uniform or p-curve, trim-and-fill.
- **Level 3:** a single-site direct replication with N of at least 500.

Classic studies were extracted only as the comparison point for the shrinkage ratio. Narrative reviews, single small studies, books, press and vendor reports were excluded.

Retraction check: 133 DOIs (every included source and every classic study) were checked for Crossref notices that update them; 0 requests failed. Retractions or expressions of concern: 0. Corrections: 3 (The effectiveness of nudging: A meta-analysis of choice architecture i, 10.1073/pnas.2107346118; A Room with a Viewpoint Revisited: Descriptive Norms and Hotel Guests', 10.1371/journal.pone.0104086; Replicating the Disease framing problem during the 2020 COVID-19 pande, 10.1371/journal.pone.0257151). The reviewer notes for these rows already cover the corrections, and none of them can change a grade: two are a meta-analysis and a replication for social proof, which is graded on its Level 1 tests, and one is a risky-choice framing study, a context version.

---

## 6. Extraction and review

Six extraction agents each took two effects and recorded, for every result: effect, tactic or context version, design, evidence level, N, k, sites, population, setting, outcome, metric, estimate, 95% CI, bias-adjusted estimate and method, heterogeneity, the location of each number (table, page or section) and a supporting quote. A separate reviewer agent per group, which did not do the extraction, reopened every source and marked each row verified, corrected or unverifiable, and listed possible omissions. Omissions that could change a grade were extracted in follow-up rounds and reviewed the same way.

- Rows extracted: 189; verified or corrected: 180; dropped as unverifiable: 9.
- Rows that set a grade or best estimate: 68 (Level 1: 27, Level 2: 36, Level 3: 5), plus 7 classic comparison rows, from 73 sources. Of the grade-setting rows, 8 were checked against the abstract only.
- Primary rows with only a point estimate (no CI, test statistic or sourced call) were reported but not graded: 6.
- Rows dropped as unverifiable: A40 Rhodes, Shulman & McClaran 2020; A41 Rhodes, Shulman & McClaran 2020; C13 Fern, Monroe & Avila 1986; C25 Fern, Monroe & Avila 1986; C28 Freedman & Fraser 1966; D24 Meyerowitz & Chaiken 1987; D26 Langer, Blank & Chanowitz 1978; D27 Langer, Blank & Chanowitz 1978 (big favor); E27 Vohs et al. 2021.
- Positive/null calls for rows without a CI come from a fixed list of row calls, each with a quoted basis from the source.

---

## 7. Effect sizes

Estimates are kept in their native metric and converted to Cohen's d where valid:  $r$  to  $d = 2r / \sqrt{1 - r^2}$ ; odds ratios to  $d = \ln(OR) \times \sqrt{3} / \pi$ ; two proportions through the odds ratio (Woolf standard error);  $t$  or  $F(1, df)$  inside a design to  $d = t \times \sqrt{1/n_1 + 1/n_2}$  with the two compared cells; Hedges'  $g$  is treated as  $d$ . CI endpoints are converted the same way. Risk ratios, percentage-point differences without counts and other native metrics are not converted; such rows count toward a grade but cannot be a best estimate. Two or more Level 1 tests in  $d$  are pooled by DerSimonian-Laird random effects.

---

## 8. Grading rule, best estimate and shrinkage ratio

A test is positive when its 95% CI lies entirely above the null (0 for differences, 1 for ratios) in the classic direction, and null otherwise. Only reviewed primary rows of the tactic version count. The first branch that fits sets the grade:

1. Level 1 tests exist: all positive gives Replicated; all null gives Failed, unless a Level 2 bias-adjusted estimate is positive (then Mixed); a mix gives Mixed.
2. No Level 1 but Level 2: a positive bias-adjusted estimate gives Replicated, a null one Failed, disagreeing methods Mixed; no bias-adjusted estimate gives Mixed.
3. Only Level 3: the grade follows the replication.
4. Nothing at Levels 1-3: Not tested at scale.

The best estimate is the Level 1 estimate in  $d$  (pooled when there are two or more); otherwise the Level 2 bias-adjusted estimate from the meta-analysis with the largest  $k$ ; otherwise the naive Level 2 estimate with the largest  $k$  (flagged); otherwise the Level 3 replication. Size labels:  $|d|$  below 0.20 very small, 0.20 to 0.49 small, 0.50 to 0.79 medium, 0.80 or more large. The shrinkage ratio is the best estimate divided by the classic study's  $d$ . The comparison study is the first classic protocol section 1 names for the tactic version that has a verified  $d$ , then the next named for the tactic version. The headline is the median ratio across effects that have one.

## 9. B2B transfer rating

High: evidence includes professionals or high-stakes field decisions with real money, and the mechanism does not depend on low attention. Medium: field evidence with consumers making real choices, or a mechanism that plausibly survives deliberate decisions, but no evidence from professional buyers. Low: only lab studies with hypothetical choices, or a mechanism that deliberation, expertise, buying committees or long sales cycles would plausibly weaken. The rating is a judgment, not a test.

## 10. Results

Of the 12 effects, 4 graded Replicated, 6 Mixed, 1 Failed and 1 Not tested at scale. 7 effects have a shrinkage ratio; the median is 0.25. 6 of the 12 predicted grades were right, and none of the three falsifiers fired.

| # | Effect                           | Grade      | Rule branch                                                      | Level 1 tests (positive / null) | Best d [95% CI]        | Basis                                  | Classic d                                            | Ratio  | B2B    |
|---|----------------------------------|------------|------------------------------------------------------------------|---------------------------------|------------------------|----------------------------------------|------------------------------------------------------|--------|--------|
| 1 | Scarcity                         | Mixed      | 2: Level 2 only, no bias-adjusted estimate                       | 0 (0 / 0)                       | 0.278 [0.224, 0.321]   | Level 2 naive (not bias-adjusted)      | 1.107 (Worchel, Lee & Adewole 1975)                  | 0.251  | Low    |
| 2 | Social proof                     | Mixed      | 1: Level 1 tests disagree                                        | 12 (7 / 5)                      | 0.125 [0.099, 0.152]   | Level 1 test                           | 0.181 (Goldstein, Cialdini & Griskevicius 2008)      | 0.693  | Medium |
| 3 | Anchoring                        | Replicated | 1: all Level 1 tests positive                                    | 1 (1 / 0)                       | 1.170 [1.110, 1.230]   | Level 1 test                           | 0.930 (Jacowitz & Kahneman 1995)                     | 1.258  | High   |
| 4 | Decoy effect                     | Failed     | 1: all Level 1 tests null                                        | 1 (0 / 1)                       | 0.003 [-0.021, 0.028]  | Level 1 test                           | none                                                 |        | Low    |
| 5 | Reciprocity                      | Replicated | 1: all Level 1 tests positive                                    | 2 (2 / 0)                       | 0.244 [-0.012, 0.500]  | random-effects pool of 2 Level 1 tests | 0.408 (Cialdini et al. 1975)                         | 0.597  | Medium |
| 6 | Foot-in-the-door                 | Mixed      | 2: Level 2 only, no bias-adjusted estimate                       | 0 (0 / 0)                       | 0.440 [0.290, 0.590]   | Level 2 naive (not bias-adjusted)      | none                                                 |        | Low    |
| 7 | Loss-aversion framing            | Mixed      | 1: Level 1 tests disagree                                        | 3 (1 / 2)                       | 0.023 [-0.006, 0.053]  | random-effects pool of 3 Level 1 tests | none                                                 |        | Medium |
| 8 | Priming                          | Mixed      | 1: Level 1 null but a Level 2 bias-adjusted estimate is positive | 2 (0 / 2)                       | -0.040 [-0.120, 0.040] | Level 1 test                           | 1.044 (Bargh, Chen & Burrows 1996)                   | -0.038 | Low    |
| 9 | Ego depletion / decision fatigue | Mixed      | 1: Level 1 tests disagree                                        | 3 (1 / 2)                       | 0.076 [0.006, 0.146]   | random-effects pool of 2 Level 1 tests | 1.785 (Baumeister, Bratslavsky, Muraven & Tice 1998) | 0.043  | Medium |

| #  | Effect                  | Grade               | Rule branch                   | Level 1 tests (positive / null) | Best d [95% CI]      | Basis        | Classic d                     | Ratio | B2B    |
|----|-------------------------|---------------------|-------------------------------|---------------------------------|----------------------|--------------|-------------------------------|-------|--------|
| 10 | Choice overload         | Replicated          | 1: all Level 1 tests positive | 2 (2 / 0)                       | 0.103 [0.092, 0.115] | Level 1 test | 1.492 (Iyengar & Lepper 2000) | 0.069 | Medium |
| 11 | "Because" justification | Not tested at scale | 4: no Level 1-3 source        | 0 (0 / 0)                       | none                 |              | none                          |       | Low    |
| 12 | Name personalization    | Replicated          | 1: all Level 1 tests positive | 1 (1 / 0)                       | 0.144 [0.081, 0.206] | Level 1 test | none                          |       | Medium |

Level 1 tests behind each grade:

- Social proof: Hallsworth et al. 2016 (positive); Gold et al. 2021 (null); Sacarny et al. 2018 (positive); Sallis et al. 2018 (positive); Dur et al. 2021 (null); Çat et al. 2024 (null); Shen, Sood & Weitzel 2026 (null); Çoker, Pechey & Jebb 2025 (null); Hallsworth et al. 2017 (positive); Muchnik, Aral & Taylor 2013 (positive); Ferraro & Price 2013 (positive); Allcott 2015 (positive)
- Anchoring: Klein et al. 2014 (positive)
- Decoy effect: Rafai et al. 2022 (null)
- Reciprocity: Genschow et al. 2021 (positive); Becker, Messer & Wolter 2012 (positive)
- Loss-aversion framing: Sallis, Harper & Sanders 2018 (positive); Dorison, Lerner, Heller et al. 2022 (null); Gold, Durlak, Sanders et al. 2019 (null)
- Priming: O'Donnell et al. 2018 (null); Verschuere et al. 2018 (null)
- Ego depletion / decision fatigue: Hagger et al. 2016 (null); Dang et al. 2021 (positive); Keller & Kiss 2021 (null)
- Choice overload: Long et al. 2025 (positive); Bertrand et al. 2010 (positive)
- Name personalization: Joinson & Reips 2007 (Study 1) (positive)

## Per-effect summaries (reviewer-checked text)

**1. Scarcity (Mixed).** Two meta-analyses find that scarcity messages raise stated purchase intentions by a small amount ( $d$  about 0.3). Neither reports a bias-adjusted estimate (only a fail-safe  $N$  and a funnel plot). In Barton et al. 2022, only 24 of 416 effect sizes measured actual behavior and only 6 came from field studies, and most purchase-intention effects came from student samples; Ladeira et al. 2023 do not report sample types. We found no preregistered multi-site replication and no large field experiment with usable statistics that tests scarcity as a sales message. The one very large field test found (Instacart, abstract only) saw shoppers buy items flagged as low in stock 25% less, not more, but the authors frame that flag as a stockout warning, so it is not a clean test of a scarcity sales message. So the effect is not shown to be false, but it has not been tested at scale in the way sales teams use it.

*B2B transfer (Low):*

The pooled evidence is almost all hypothetical purchase intentions, mostly from students (Barton: 178 of 193 purchase-intention effects hypothetical, 152 from student samples, only 6 field effect sizes in the full data set), there is no evidence with professionals or organizational buyers, and the one large real-purchase test (Instacart, abstract only) went the other way, although its low-stock flag was a stockout warning rather than a sales message. The sources do not test whether deliberation or buying committees weaken the effect; in hypothetical scenarios, Barton et al. found effects for high-involvement products no smaller than for low-involvement ones.

**2. Social proof (Mixed).** Telling people what most others do can move real behavior in large field tests, but only sometimes and by small amounts. Large or preregistered trials of tax reminder letters (+1.3 percentage points in payment within 23 days, no longer detectable by day 48), letters to high-prescribing GP practices (3.3% fewer

antibiotic prescriptions), organ donor sign-up prompts (2.3% to 2.8% sign-up) and Medicare prescriber letters (11.1% fewer quetiapine days, in letters that also said prescribing was under review) found positive effects. Energy reports and a water letter that compare a household with its neighbours cut use by about 1 to 5 percent (Opower first-year average 1.31% across 111 trials; a strong-norm water letter, which also carried advice and a prosocial appeal, about 4.8%), and predictions from the first utilities overstated what later ones achieved. Other large or preregistered trials with bank savings, vaccination ads, software downloads, cafeteria meals and a second GP letter found no increase, and the vaccination ad backfired in Taiwan. The bias-corrected meta-analytic estimates closest to the tactic are near zero (health messaging  $d = 0.01$ ; Maier et al.'s broader information-nudge category  $d = 0.00$ ), while broader social-influence and social-comparison estimates stay small and positive after trim-and-fill ( $g = 0.16$  and  $0.15$ ). The German replication of the hotel towel study found no advantage for the norm message (81.9% vs 83.7% reuse). The fair reading is a small effect that depends on context, not a reliable lever, and the null results do not prove it never works.

#### *B2B transfer (Medium):*

There are large real-choice field tests with professionals (GP practices, Medicare prescribers, meta-analyses of healthcare worker trials), but the professional evidence is peer-comparison feedback sent by an authority (the Medicare letters also said prescribing was under review), with mixed trial results (two positive, one null) and very small pooled effects (social comparison alone SMD 0.05, fixed-effect, Tang 2021). The one test of a pure popularity count with software users (inflated PyPI download counts) found no effect, but the paper does not show that users saw the counts, and none of the evidence involves organizational buyers.

**3. Anchoring (Replicated).** Anchoring holds up in large tests: all four anchoring items in the preregistered Many Labs 1 replication produced large effects ( $d = 1.17$  to  $2.42$ ), numerically above the original estimate ( $d = 0.93$ ), although Many Labs 1 notes that the original used a different, test-retest design, so that comparison is not like for like. The largest meta-analysis (1,280 effect sizes) finds  $g = 0.83$ ; every publication-bias correction it reports stays clearly above zero, but the corrected values range from about 0.45 to 0.90 and the authors say the effect may shrink by as much as half. The effect is smaller in the forms closest to pricing. Anchors on willingness to pay average about  $r = 0.27$  ( $r = 0.21$  for random anchors), and the five incentivized results that used transparently random anchors show no effect ( $g = 0.03$ , too few to grade). In 16 pay-what-you-want field studies with real payments, anchoring appeared only under some conditions, whereas hypothetical versions produced large effects; that paper's results are known only from its abstract. First offers pull negotiated agreements toward them ( $g = 1.14$ ), but that estimate is not robust to every bias correction, rests mostly on vignette, lab and classroom studies, and ambitious offers also raise the risk of no deal. A number that appears only in a restaurant or product name did not move judgments in Many Labs 2 or in a preregistered study of 874 people, even though standard anchoring was large in the same sample. The defensible claim for sales teams is that a plausible, relevant number such as an opening offer shifts the other side's estimates, not that any arbitrary number does.

#### *B2B transfer (High):*

Anchoring has evidence with professionals (legal professionals and real-estate agents in single studies, and a legal meta-analysis, all checked at abstract level) and from real-money settings (pay-what-you-want field studies and eBay negotiations, not graded), and experience did not remove it in the legal study; the usable form in B2B is a plausible, relevant anchor such as a list price or first offer, because transparently random anchors show no effect on incentivized valuations and the first-offer evidence comes mostly from vignette, lab and classroom studies.

**4. Decoy effect (Failed).** The decoy effect is reliable in hypothetical online choice tasks: in a 2025 registered report with 2,293 participants, people chose the dominating option in about 60% of target-or-competitor choices, against 50% expected with no effect, with a larger effect for numeric than graphical displays. That measure is the target's share relative to its competitor, not its overall share. Outside the lab the evidence is weak. The one large field experiment, on a flight booking site with more than 140,000 search sessions, found no shift toward the targeted option, and studies that show attributes as pictures or let people experience the product often find nothing; the evidence for verbal descriptions conflicts. The newest meta-analysis, an unreviewed preprint not used for the grade, finds in its robust Bayesian sensitivity analysis that the gain in the target's overall choice share is about zero once

small-study and publication bias are modeled, while its standing relative to the competitor survives. For B2B buyers who deliberate, compare options in detail and decide in groups, a decoy option is unlikely to be a dependable lever, although this does not prove it never works.

*B2B transfer (Low):*

The large real-choice test found nothing, the effect is small or absent when attributes are pictorial or experienced rather than numeric (Frederick, Lee & Baskin 2014, abstract; Cai & Pleskac 2026 preprint, Table 9), and in incentivized objective-value tasks it fades when people have time to revise their choice (Crosetto & Gaudeul 2024, abstract), which resembles deliberate committee buying.

**5. Reciprocity (Replicated).** Giving first raises compliance in large tests, but the effects are small and can fade. In a Swiss randomized field experiment (2,437 voucher recipients and 17,234 controls, about 17,300 still reachable at follow-up), an unsolicited adult-training voucher raised participation in the next survey round by about 5 points (71.9% vs 66.6%), with no effect 18 months later. The one preregistered field replication of door-in-the-face found here (Genschow et al. 2021, N = 391) reproduced the 1975 pattern, 51% vs 34% agreement. Meta-analyses put door-in-the-face at about  $r = .10$  to  $.14$  for overall or verbal compliance and  $r = .05$  to  $.08$  for behavioral follow-through (the  $r = .05$  estimate was not statistically significant). Sending a survey incentive with the request rather than promising it raised the odds of response by about half (OR 1.53, 35 trials, I<sup>2</sup> = 89%), with no bias-adjusted estimate reported. Cross-study moderator analyses suggest door-in-the-face works more reliably when the same person makes both requests, there is no delay between them and the request serves a prosocial cause. In the voucher experiment only the larger vouchers had significant effects, and recipients who used their voucher responded more while non-users did not, a comparison that is not randomized. No test isolated professional buyers. One preregistered field experiment with 1,950 non-profit and educational organizations, available only as a preprint and excluded from grading, found that adding a \$1.20 branded planner to a cold mailing lowered response (14.9% vs 20.3% scanned the card's QR code), which is a caution for B2B use rather than proof the effect is false.

*B2B transfer (Medium):*

Large field tests with real behavior (survey participation, verbal agreement to volunteer) support the effect, but no source isolates professional buyers; the gift effect was gone by 18 months, and door-in-the-face effects were weaker or reversed with a different requester or a delay between requests. The evidence involving executives or organizations leans the other way but is weak: in between-study meta-analyses of published surveys, the usual response-raising techniques were not associated with higher executive response (Cycyota & Harrison 2006, abstract) and incentive use had a more negative slope for top executives (Anseel et al. 2010), and one preregistered organizational field test (a preprint, excluded from grading) found a \$1.20 gift lowered response, so Medium is borderline with Low.

**6. Foot-in-the-door (Mixed).** No large preregistered replication of foot-in-the-door was found, so the evidence is older meta-analyses of mostly small studies plus one 2024 synthesis. They find a positive but small effect: about  $r = .13$  to  $.17$  (d about 0.25 to 0.35) in the 1980s meta-analyses, and  $g = 0.44$  (95% CI 0.29 to 0.59) in the 2024 synthesis for prosocial requests only, with no bias-adjusted estimate available to this review. That is below the Freedman & Fraser 1966 result of 53% vs 22% compliance (d about 0.75). Nearly half the studies in the 1983 review found no effect or a reversal, and the effect appears mainly when the requests serve a prosocial cause. One direct replication in Poland and Ukraine found no effect in either country in its 2013 wave (abstract only; its sample size and registration status were not verified). The fair reading is a small effect that is unproven at scale, not an effect proven false.

*B2B transfer (Low):*

The evidence comes mostly from small experiments with private individuals and students; Dillard et al. 1984 and Ferguson et al. 2024 both find the effect depends on a prosocial request, which a sales ask usually is not, and no test with professional buyers was found.

**7. Loss-aversion framing (Mixed).** The playbook rule that a message framed as a loss avoided beats the same message framed as a gain obtained does not hold up as a general rule. The three large tests disagree. A UK web trial with over a million visitors found a tiny loss-frame edge in organ-donor sign-ups (3.0% vs 2.8% among the 273,491 people shown those two messages), but its loss and gain messages differed in content as well as frame, as the authors note. A preregistered 84-country study (15,929 people) and an 11,038-person NHS invitation trial found no reliable difference. Meta-analyses of 165 message comparisons pool to essentially zero ( $r = .016$  favoring gain frames, CI includes zero), with small topic differences that sometimes favor gain frames, and none of the message-framing meta-analyses reports a bias-adjusted estimate. A Cochrane review of 16 health-behavior studies and a meta-analysis of 40 charity-appeal comparisons also find little or no difference. People do weigh losses more than gains in risky choices, and risky-choice framing replicates strongly, but in these tests that has not turned into a reliable persuasion advantage for loss-framed messages.

*B2B transfer (Medium):*

Field tests with members of the public making real choices exist (organ-donor sign-ups, health-check uptake), but none with professionals or organizational buyers, and in those field tests the loss-vs-gain difference was a fraction of a percentage point.

**8. Priming (Mixed).** In the large preregistered multi-lab replications, priming people with a concept produced no detectable change in their behaviour. This holds for professor priming and trivia scores (23 labs, 0.14 percentage points, 95% CI [-0.71, 1.00]) and for moral reminders and cheating (19 labs,  $d = -0.04$  [-0.12, 0.04]). A large single-site replication with 620 commercial bankers, which was not preregistered, also found no significant effect of professional-identity priming on cheating ( $d = 0.06$ ). Multi-site tests of priming on attitudes and judgments (money, flag, heat, hostile words) also land at or near zero, for example  $d = -0.02$  for money images. Meta-analyses of the existing literature, mostly lab studies, find a small-to-moderate average ( $d$  about 0.31 to 0.37). Bias adjustments disagree. Some leave a small positive effect ( $d = 0.14$  to 0.31 in the two largest behavioural-priming meta-analyses), while others are close to zero, including the preregistered studies within the largest meta-analysis ( $d = 0.02$  [-0.20, 0.24], 13 effects). So a reliable behavioural priming effect is not shown, though small effects in specific settings are not ruled out.

*B2B transfer (Low):*

Large preregistered multi-lab tests with behavioural outcomes find effects near zero, and the one large replication with professionals (620 commercial bankers) found no significant effect. The meta-analysed effects come from incidentally presented primes, and deliberate committee buying would plausibly weaken that mechanism.

**9. Ego depletion / decision fatigue (Mixed).** Preregistered multi-lab tests of the sequential-task ego-depletion paradigm find at most a small effect. Two found no significant effect:  $d = 0.04$  [-0.07, 0.15] across 23 labs, and  $d = 0.06$  across 36 labs, reported as nonsignificant. One found  $d = 0.10$  [0.01, 0.19] across 12 labs. All three are far below the original radish-and-chocolate result ( $t(44) = 6.03$  with about 21 to 25 people per group). A preregistered field experiment with 1,143 schoolchildren found 4.4 percentage points more cheating on its primary outcome, but its 95% CI includes zero. Bias-corrected meta-analyses of earlier studies shrink the average ( $d$  or  $g$  about 0.4 to 0.7 before correction) to roughly zero by PET-PEESE, though trim-and-fill still leaves about 0.24. A newer multi-lab study with a much longer depleting task reports  $d = 0.35$ , but most of its samples were not preregistered. The 'decision fatigue' patterns among judges and doctors are observational, and the judge finding is contested, so effects on real professional decisions are untested by experiment rather than disproven.

*B2B transfer (Medium):*

Fatigue built up over many sequential decisions plausibly affects deliberate professional work, and observational data from judges and clinicians are consistent with that, though the judge finding is contested. But there is no randomized test with professionals or buyers, and the preregistered experimental effect is at most small ( $d$  about 0.1 in multi-lab tests).

**10. Choice overload (Replicated).** Large tests do not support the jam study at anything like its original size: 30% vs 3% buying among tasters in 2000 compares with differences of about two percentage points or less in big field experiments. Showing one example loan instead of four in a mailer raised loan applications by 0.68 points on an 8.5% base, and recommending two rather than four products raised same-day purchases from 4.29% to 5.13%, but in the same experiment one recommended product sold worse than four (3.08% vs 4.29%). A supermarket experiment that cut shelf variety (35,694 passers-by, 3,291 of whom stopped) found no significant effect on purchase, with a wide interval. The two meta-analyses disagree: one finds a mean near zero across 63 conditions, the other a small mean effect ( $d = 0.17$ ) after excluding studies where people chose between assortments, and neither reports a bias-corrected estimate. The fair reading is that fewer options can help when choices are complex and buyers are unsure, not that more choice reliably hurts sales.

*B2B transfer (Medium):*

The field evidence involves consumers making real purchases and real loan decisions, but no study tests professional or organizational buyers, and the Scheibehenne et al. 2010 meta-regression finds that expertise or prior preferences reduce or reverse the effect, which suggests weaker effects for experienced buyers.

**11. "Because" justification (Not tested at scale).** No large replication, multi-site project or large field experiment of the 'because' effect was found, so it has not been tested at scale; that is different from being shown false. The evidence for the tactic is the small 1978 copy-machine study and a few small follow-ups; this review could not open a legitimate full-text copy of the 1978 paper, so its compliance figures are not reported here. The one qualifying meta-analysis is about real arguments, not empty reasons: spelling out the supporting argument in a persuasive message raises persuasion a little ( $r = .116$ , 95% CI .052 to .180,  $k = 27$ ; O'Keefe 1998). That fits with giving genuine reasons, but it does not test whether filler reasons work.

*B2B transfer (Low):*

No test with professionals, organizational buyers or high-stakes decisions exists; the evidence is one small classic study plus a meta-analysis of real rather than empty arguments, and an effect described as a low-attention response would plausibly weaken under deliberate, committee-based B2B decisions.

**12. Name personalization (Replicated).** Adding the recipient's name tends to raise response, but by small amounts. In a 10,000-person email experiment, using the student's name raised panel sign-ups from about 12.5% to 15.6% of invitations, and the best-known marketing experiment reports (in its abstract) email opens rising from 9.05% to 10.80% with the name in the subject line. Across 12 randomized trials of email survey invitations, most of which put the name in the salutation, personalization raised the odds of a response by about a quarter, though no bias-corrected estimate exists. A 2023 replication with 3,521 university students emailed in German found no first-name subject-line effect; its authors suggest the trick may have worn off as it became common, but the study cannot separate that from differences in audience and country.

*B2B transfer (Medium):*

Field tests use real inboxes and real decisions to respond, and two samples are closer to professional audiences (early-career scientists and engineers, and local politicians), but neither of those tests is a qualifying design, the one direct replication of the subject-line version failed, and every outcome is an open, sign-up or survey response rather than a purchase.

## Sensitivity (not pre-registered)

| Scenario | What changes                                                                                                            | Replicated / Mixed / Failed / Not tested | Median ratio | Grades that move |
|----------|-------------------------------------------------------------------------------------------------------------------------|------------------------------------------|--------------|------------------|
| S0       | Main analysis as pre-registered                                                                                         | 4 / 6 / 1 / 1                            | 0.25         | none             |
| S1       | Jung, Perfecto & Nelson 2016 (B27) counted as a Level 1 field test and called null, as the working-paper abstract reads | 3 / 7 / 1 / 1                            | 0.25         | Anchoring        |

| Scenario | What changes                                                                                                                     | Replicated / Mixed / Failed / Not tested | Median ratio | Grades that move                                                        |
|----------|----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|--------------|-------------------------------------------------------------------------|
| S2       | Ong 2021 (F08) counted as Level 1, taking the 35,694 passers-by as the units                                                     | 3 / 7 / 1 / 1                            | 0.25         | Choice overload                                                         |
| S3       | Long et al. 2025 primary contrast taken as 1 vs 4 products (F06) instead of 2 vs 4 (F05)                                         | 3 / 7 / 1 / 1                            | 0.25         | Choice overload                                                         |
| S4       | Bertrand et al. 2010 (F07) left out: one of 13 mailer features tested at once                                                    | 4 / 6 / 1 / 1                            | 0.25         | none                                                                    |
| S5       | Iyengar & Lepper 2000 classic d from the phi coefficient of its 2x2 table instead of the odds ratio                              | 4 / 6 / 1 / 1                            | 0.25         | none                                                                    |
| S6       | Muchnik, Aral & Taylor 2013 (A42) left out, as the reviewer set it                                                               | 4 / 6 / 1 / 1                            | 0.25         | none                                                                    |
| S7       | Every flagged call against the effects at once (S1, S2, S3, S4, S9 and S10)                                                      | 1 / 7 / 3 / 1                            | 0.25         | Anchoring; Loss-aversion framing; Choice overload; Name personalization |
| S8       | Loss-aversion framing compared with Tversky & Kahneman 1981, the protocol's risky-choice (context) classic                       | 4 / 6 / 1 / 1                            | 0.16         | none                                                                    |
| S9       | Sallis, Harper & Sanders 2018 loss-vs-gain test (D01) left out: its two letters differ in content as well as frame               | 4 / 5 / 2 / 1                            | 0.25         | Loss-aversion framing                                                   |
| S10      | Joinson & Reips 2007 Study 1 (F11) below Level 1, counting the 9,294 delivered emails instead of 10,000 invitations              | 3 / 7 / 1 / 1                            | 0.25         | Name personalization                                                    |
| S11      | Worchel, Lee & Adewole 1975 classic d from 2 $\sqrt{F} / \sqrt{122}$ , the full design's error df, instead of the two cell sizes | 4 / 6 / 1 / 1                            | 0.42         | none                                                                    |

## 11. Verification

An independent re-analysis, written from the study plan, the protocol and the extract alone without seeing our code, recomputed every grade, count, best estimate, classic d, shrinkage ratio and headline number. After one rule clarification (the classic-study fallback stops at classics listed for the tactic version; see section 13), 120 of 120 compared items match. Its first pass also found that our code counted point-estimate-only Level 2 rows, which our call rule excludes; the code was fixed and no grade moved.

## 12. Limitations

- Publication bias: only 11 of the 36 grade-setting meta-analytic rows report a bias-adjusted estimate. Scarcity and foot-in-the-door rest on naive meta-analytic estimates, which are likely too high.
- Comparison studies: several classics used designs or metrics that differ from the replications (anchoring's original was test-retest; Worchel et al. 1975 reports an F from a 12-cell design; Iyengar & Lepper 2000 conditions on stopping at the display). The median ratio moves from 0.25 to between 0.16 and 0.42 under the alternatives in scenarios S8 and S11.
- Single tests: the decoy effect (Failed) and name personalization (Replicated) each rest on one Level 1 test; anchoring, choice overload and loss framing each turn on one judgment call (scenarios S1, S2, S3, S9, S10). With every flagged call against the effects, 1 effect replicates.

- Access: 8 grade-setting rows were checked against abstracts only, and several sources were read in author, repository, SSRN or NBER versions instead of the typeset article (noted by the reviewers in the evidence table). Unverifiable numbers were dropped. One open-access social-proof paper (Doskeland et al. 2025, Journal of Business Research; two natural field experiments with 455,509 bank customers) was excluded because its full text could not be downloaded and its abstract gives no effect size against control; social proof already has both positive and null Level 1 tests, so one more test cannot change its Mixed grade, but its estimate is missing from the evidence.
- Choice overload's best estimate (Long et al. 2025) is computed from the purchase rates the reviewer verified and from per-arm sizes in an appendix the reviewer could not open. The arm sizes sum to the verified total of 1,615,325 and affect only the width of the CI, not the point estimate. Sahni et al. 2018 (not graded) is marked verified but was checked against its abstract only.
- Search coverage: the Semantic Scholar pass stopped at 44 of 125 queries; OpenAlex, PsycINFO, Web of Science and Scopus were unavailable; screening was done by language-model agents. Known-item lookups, targeted searches, citation chasing and reviewer omission checks reduce but do not remove the risk of missed tests.
- Omissions listed but not extracted: the group E reviewer suggested 11 sources for priming and ego depletion. They were not extracted because neither grade can change under the rule. They are listed in section 14.
- Population: nearly all graded evidence comes from consumers, students, survey respondents or online panels, not professional buyers.
- Raw search storage: the protocol planned to store the raw search results on Ken's servers. The servers could not be reached at the time, so the raw search results were archived separately, with a checksum recorded, and nothing was written to the servers.

---

## 13. Deviations from the pre-registered protocol

- OpenAlex could not be searched (no API key). The Semantic Scholar pass was stopped at 44 of 125 queries.
- Raw search results were archived separately instead of on Ken's servers (see section 12).
- Operational rules written after the search, each logged with its reason in the study plan: which rows count, the null value per metric, the best-estimate fallback order, the classic-study fallback order (clarified after verification to classics listed for the tactic version), the row-call rule for results without a CI, dropping numbers read only from search-engine excerpts, and the t-to-d formula for contrasts inside larger designs.
- The sensitivity scenarios (section 10) were added after extraction and are reported next to the pre-registered analysis, not in place of it.

---

## 14. Sources

Sources behind every grade, best estimate and classic comparison (the study recorded the full set of 132 included sources, each with its reason).

### Anchoring

- Jacowitz, K.; Kahneman, D. (1995). Measures of Anchoring in Estimation Tasks. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/01461672952111004>. [classic; checked against fulltext]
- Klein, R.; Ratliff, K.; Vianello, M.; Adams, R.; Bahník, Š.; Bernstein, M.; et al. (2014). Investigating Variation in Replicability. *Social Psychology*. <https://doi.org/10.1027/1864-9335/a000178>. [Level 1; checked against fulltext]
- Bystranowski, P.; Janik, B.; Próchnicki, M.; Skórska, P. (2021). Anchoring effect in legal decision-making: A meta-analysis.. *Law and Human Behavior*. <https://doi.org/10.1037/lhb0000438>. [Level 2; checked against abstract]

- Ioannidis, K.; Offerman, T.; Sloof, R. (2020). On the effect of anchoring on valuations when the anchor is transparently uninformative. *Journal of the Economic Science Association*. <https://doi.org/10.1007/s40881-020-00094-1>. [Level 2; checked against fulltext]
- Li, L.; Maniadis, Z.; Sedikides, C. (2021). Anchoring in Economics: A Meta-Analysis of Studies on Willingness-To-Pay and Willingness-To-Accept. *Journal of Behavioral and Experimental Economics*. <https://doi.org/10.1016/j.socec.2020.101629>. [Level 2; checked against fulltext]
- Petrowsky, H.; Boecker, L.; Escher, Y.; Frech, M.; Frieze, M.; Galinsky, A.; et al. (2025). The power and peril of first offers in negotiations: a conceptual, meta-analytic, and experimental synthesis. *Organizational Behavior and Human Decision Processes*. <https://doi.org/10.1016/j.obhdp.2025.104448>. [Level 2; checked against fulltext]
- Schley, D.; Weingarten, E. (2026). Fifty Years of Anchoring Effects: A Theoretical Reintegration and Meta-Analysis. *Management Science*. <https://doi.org/10.1287/mnsc.2023.03238>. [Level 2; checked against fulltext]
- Yoon, S.; Fong, N.; Dimoka, A. (2019). The robustness of anchoring effects on preferential judgments. *Judgment and Decision Making*. <https://doi.org/10.1017/s1930297500006148>. [Level 2; checked against fulltext]
- Shanks, D.; Barbieri-Hermitte, P.; Vaddillo, M. (2020). Do Incidental Environmental Anchors Bias Consumers' Price Estimations?. *Collabra: Psychology*. <https://doi.org/10.1525/collabra.310>. [Level 3; checked against fulltext]

### Choice overload

- Iyengar, Sheena S.; Lepper, Mark R. (2000). When choice is demotivating: Can one desire too much of a good thing?. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.79.6.995>. [classic; checked against fulltext]
- Bertrand, Marianne; Karlan, Dean; Mullainathan, Sendhil; Shafir, Eldar; Zinman, Jonathan (2010). What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. *Quarterly Journal of Economics*. <https://doi.org/10.1162/qjec.2010.125.1.263>. [Level 1; checked against fulltext]
- Long, Xiaoyang; Sun, Jiankun; Dai, Hengchen; Zhang, Dennis; Zhang, Jianfeng; Chen, Yujie et al. (2025). The Choice Overload Effect in Online Recommender Systems. *Manufacturing & Service Operations Management*. <https://doi.org/10.1287/msom.2022.0659>. [Level 1; checked against fulltext]
- Chernev, Alexander; Böckenholt, Ulf; Goodman, Joseph (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*. <https://doi.org/10.1016/j.jcps.2014.08.002>. [Level 2; checked against fulltext]
- Scheibehenne, Benjamin; Greifeneder, Rainer; Todd, Peter M. (2010). Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload. *Journal of Consumer Research*. <https://doi.org/10.1086/651235>. [Level 2; checked against fulltext]

### Decoy effect

- Rafai, I.; Babutsidze, Z.; Delahaye, T.; Hanaki, N.; Acuna-Agost, R. (2022). No evidence of attraction effect among recommended options: A large-scale field experiment on an online flight aggregator. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2021.113672>. [Level 1; checked against fulltext]

### Ego depletion / decision fatigue

- Roy F. Baumeister, Ellen Bratslavsky, Mark Muraven, Dianne M. Tice (1998). Ego depletion: Is the active self a limited resource?. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.74.5.1252>. [classic; checked against fulltext]
- Junhua Dang, Paul Barker, Anna Baumert, Margriet Bentvelzen, Elliot Berkman, Nita Buchholz et al. (2021). A Multilab Replication of the Ego Depletion Effect. *Social Psychological and Personality Science*. <https://doi.org/10.1177/1948550619887702>. [Level 1; checked against fulltext]
- M. S. Hagger, N. L. D. Chatzisarantis, H. Alberts, C. O. Anggono, C. Batailler, A. R. Birt et al. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*.

<https://doi.org/10.1177/1745691616652873>. [Level 1; checked against abstract]

- Tamás Keller, Hubert János Kiss (2021). Do exhausted primary school students cheat more? A randomized field experiment. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0260141>. [Level 1; checked against fulltext]
- Evan C. Carter, Michael E. McCullough (2014). Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated?. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.00823>. [Level 2; checked against fulltext]
- Evan C. Carter, Lilly M. Kofler, Daniel E. Forster, Michael E. McCullough (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource.. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000083>. [Level 2; checked against fulltext]
- Junhua Dang (2018). An updated meta-analysis of the ego depletion effect. *Psychological Research*. <https://doi.org/10.1007/s00426-017-0862-x>. [Level 2; checked against fulltext]

### Foot-in-the-door

- Ferguson, R.; Kaufmann, L.; Brown, A.; de la Piedad Garcia, X. (2024). Influences of past moral behavior on future behavior: A review of sequential moral behavior studies using meta-analytic techniques.. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000441>. [Level 2; checked against abstract]

### Loss-aversion framing

- Dorison, Charles A.; Lerner, Jennifer S.; Heller, Blake H.; Rothman, Alexander J.; Kawachi, Ichiro I.; Wang, Ke; et al. (2022). In COVID-19 Health Messaging, Loss Framing Increases Anxiety with Little-to-No Concomitant Benefits: Experimental Evidence from 84 Countries. *Affective Science*. <https://doi.org/10.1007/s42761-022-00128-3>. [Level 1; checked against fulltext]
- Gold, Natalie; Durlak, Caroline; Sanders, Jet G.; Thompson, Katherine; Chadborn, Tim (2019). Applying behavioural science to increase uptake of the NHS Health Check: a randomised controlled trial of gain- and loss-framed messaging in the national patient information leaflet. *BMC Public Health*. <https://doi.org/10.1186/s12889-019-7754-5>. [Level 1; checked against fulltext]
- Sallis, Anna; Harper, Hugo; Sanders, Michael (2018). Effect of persuasive messages on National Health Service Organ Donor Registrations: a pragmatic quasi-randomised controlled trial with one million UK road taxpayers. *Trials*. <https://doi.org/10.1186/s13063-018-2855-5>. [Level 1; checked against fulltext]
- Akl, Elie A; Oxman, Andrew D; Herrin, Jeph; Vist, Gunn E; Terrenato, Irene; Sperati, Francesca; et al. (2011). Framing of health information messages. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.cd006777.pub2>. [Level 2; checked against fulltext]
- O'Keefe, Daniel J.; Jensen, Jakob D. (2006). The Advantages of Compliance or the Disadvantages of Noncompliance? A Meta-Analytic Review of the Relative Persuasive Effectiveness of Gain-Framed and Loss-Framed Messages. *Annals of the International Communication Association (Communication Yearbook 30)*. <https://doi.org/10.1080/23808985.2006.11679054>. [Level 2; checked against fulltext]
- O'Keefe, Daniel J.; Jensen, Jakob D. (2007). The Relative Persuasiveness of Gain-Framed Loss-Framed Messages for Encouraging Disease Prevention Behaviors: A Meta-Analytic Review. *Journal of Health Communication*. <https://doi.org/10.1080/10810730701615198>. [Level 2; checked against fulltext]
- O'Keefe, Daniel J.; Jensen, Jakob D. (2009). The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Detection Behaviors: A Meta-Analytic Review. *Journal of Communication*. <https://doi.org/10.1111/j.1460-2466.2009.01417.x>. [Level 2; checked against fulltext]
- O'Keefe, Daniel J.; Nan, Xiaoli (2012). The Relative Persuasiveness of Gain- and Loss-Framed Messages for Promoting Vaccination: A Meta-Analytic Review. *Health Communication*. <https://doi.org/10.1080/10410236.2011.640974>. [Level 2; checked against fulltext]

- O'Keefe, Daniel J.; Wu, Daisy (2012). Gain-Framed Messages Do Not Motivate Sun Protection: A Meta-Analytic Review of Randomized Trials Comparing Gain-Framed and Loss-Framed Appeals for Promoting Skin Cancer Prevention. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph9062121>. [Level 2; checked against fulltext]
- Waheed, Hassam (2023). Nudging smokers away from lighting up: A meta-analysis of framing effect in current smokers. *Journal of Behavioral and Experimental Economics*. <https://doi.org/10.1016/j.socec.2023.101998>. [Level 2; checked against abstract]
- Xu, Jie; Huang, Guanxiong (2020). The relative effectiveness of gain-framed and loss-framed messages in charity advertising: Meta-analytic evidence and implications. *International Journal of Nonprofit and Voluntary Sector Marketing*. <https://doi.org/10.1002/nvsm.1675>. [Level 2; checked against abstract]

### Name personalization

- Joinson, Adam N.; Reips, Ulf-Dietrich (2007). Personalized salutation, power of sender and response rates to Web-based surveys. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2004.12.011>. [Level 1; checked against fulltext]
- Edwards, Philip James; Roberts, Ian; Clarke, Mike J; DiGuseppi, Carolyn; Woolf, Benjamin; Perkins, Chloe (2023). Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.MR000008.pub5>. [Level 2; checked against fulltext]
- Defau, Laurens; Zauner, Alexander (2023). Personalized subject lines in email marketing. *Marketing Letters*. <https://doi.org/10.1007/s11002-023-09701-7>. [Level 3; checked against fulltext]

### Priming

- John A. Bargh, Mark Chen, Lara Burrows (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action.. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.71.2.230>. [classic; checked against fulltext]
- Michael O'Donnell, Leif D. Nelson, Evi Ackermann, Balazs Aczel, Athfah Akhtar, Silvio Aldrovandi et al. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691618755704>. [Level 1; checked against fulltext]
- Bruno Verschuere, Ewout H. Meijer, Ariane Jim, Katherine Hoogesteyn, Robin Orthey, Randy J. McCarthy et al. (2018). Registered Replication Report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245918781032>. [Level 1; checked against fulltext]
- Nicola J. Buckland, Vanessa Er, Ian Redpath, Kristine Beaulieu (2018). Priming food intake with weight control cues: systematic review with a meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*. <https://doi.org/10.1186/s12966-018-0698-9>. [Level 2; checked against fulltext]
- Wenhao Dai, Tianshu Yang, Benjamin X. White, Ryan Palmer, Emily K. Sanders, Jack A. McDonald et al. (2023). Priming behavior: A meta-analysis of the effects of behavioral and nonbehavioral primes on overt behavioral outcomes.. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000374>. [Level 2; checked against fulltext]
- Gary P. Latham, Xiao Chen, Ronald F. Piccolo, Guy Itzhakov (2023). An updated meta-analysis of the primed goal-organizational behaviour relationship. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.221494>. [Level 2; checked against fulltext]
- Paul Lodder, How Hwee Ong, Raoul P. P. P. Grasman, Jelte M. Wicherts (2019). A comprehensive meta-analysis of money priming.. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000570>. [Level 2; checked against abstract]
- David R. Shanks, Miguel A. Vadillo (2021). Publication bias and low power in field studies on goal priming. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.210544>. [Level 2; checked against fulltext]

- Evan Weingarten, Qijia Chen, Maxwell McAdams, Jessica Yi, Justin Hepler, Dolores Albarracín (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words.. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000030>. [Level 2; checked against fulltext]
- Michiel van Elk, Dora Matzke, Quentin F. Gronau, Maime Guan, Joachim Vandekerckhove, Eric-Jan Wagenmakers (2015). Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2015.01365>. [Level 2; checked against fulltext]
- Gomes C, McCullough M (2015). The effects of implicit religious primes on dictator game allocations: A preregistered replication experiment.. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000027>. [Level 3; checked against fulltext]
- Zoe Rahwan, Erez Yoeli, Barbara Fasolo (2019). Heterogeneity in banker culture and its influence on dishonesty. *Nature*. <https://doi.org/10.1038/s41586-019-1741-y>. [Level 3; checked against fulltext]

## Reciprocity

- Cialdini, R.; Vincent, J.; Lewis, S.; Catalan, J.; Wheeler, D.; Darby, B. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique.. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/h0076284>. [classic; checked against fulltext]
- Becker, S.; Messer, D.; Wolter, S. (2012). A Gift is Not Always a Gift: Heterogeneity and Long-term Effects in a Gift Exchange Experiment. *Economica*. <https://doi.org/10.1111/ecca.12004>. [Level 1; checked against fulltext]
- Genschow, O.; Westfal, M.; Crusius, J.; Bartosch, L.; Feikes, K.; Pallasch, N. et al. (2021). Does social psychology persist over half a century? A direct replication of Cialdini et al.'s (1975) classic door-in-the-face technique.. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspa0000261>. [Level 1; checked against fulltext]
- Edwards, P.; Roberts, I.; Clarke, M.; DiGuseppi, C.; Woolf, B.; Perkins, C. (2023). Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.MR000008.pub5>. [Level 2; checked against fulltext]
- O'Keefe, D.; Hale, S. (2001). An odds-ratio-based meta-analysis of research on the door-in-the-face influence strategy. *Communication Reports*. <https://doi.org/10.1080/08934210109367734>. [Level 2; checked against fulltext]

## Scarcity

- Worchel, Stephen; Lee, Jerry; Adewole, Akanbi (1975). Effects of supply and demand on ratings of object value.. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.32.5.906>. [classic; checked against fulltext]
- Barton, Belinda; Zlatevska, Natalina; Oppewal, Harmen (2022). Scarcity tactics in marketing: A meta-analysis of product scarcity effects on consumer purchase intentions. *Journal of Retailing*. <https://doi.org/10.1016/j.jretai.2022.06.003>. [Level 2; checked against fulltext]
- Ladeira, Wagner Junior; Lim, Weng Marc; de Oliveira Santini, Fernando; Rasul, Tareq; Perin, Marcelo G.; Altinay, Levent (2023). A meta-analysis on the effects of product scarcity. *Psychology & Marketing*. <https://doi.org/10.1002/mar.21816>. [Level 2; checked against fulltext]

## Social proof

- Goldstein, Noah J.; Cialdini, Robert B.; Griskevicius, Vladas (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research*. <https://doi.org/10.1086/586910>. [classic; checked against fulltext]
- Allcott, Hunt (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*. <https://doi.org/10.1093/qje/qjv015>. [Level 1; checked against fulltext]
- Dur, Robert; Fleming, Dimitry; van Garderen, Marten; van Lent, Max (2021). A social norm nudge to save more: A field experiment at a retail bank. *Journal of Public Economics*. <https://doi.org/10.1016/j.jpubeco.2021.104443>. [Level 1; checked against fulltext]

- Ferraro, Paul J.; Price, Michael K. (2013). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_00344](https://doi.org/10.1162/rest_a_00344). [Level 1; checked against fulltext]
- Gold, Natalie; Ratajczak, Michael; Sallis, Anna; Saei, Ayoub; Watson, Robin; van Schaik, Paul; et al. (2021). Provision of social-norms feedback to general practices whose antibiotic prescribing is increasing: a national randomized controlled trial. *Journal of Public Health*. <https://doi.org/10.1007/s10389-021-01645-4>. [Level 1; checked against fulltext]
- Hallsworth, Michael; Chadborn, Tim; Sallis, Anna; Sanders, Michael; Berry, Daniel; Greaves, Felix; et al. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet*. [https://doi.org/10.1016/s0140-6736\(16\)00215-4](https://doi.org/10.1016/s0140-6736(16)00215-4). [Level 1; checked against fulltext]
- Hallsworth, Michael; List, John A.; Metcalfe, Robert D.; Vlaev, Ivo (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*. <https://doi.org/10.1016/j.jpubeco.2017.02.003>. [Level 1; checked against fulltext]
- Muchnik, Lev; Aral, Sinan; Taylor, Sean J. (2013). Social Influence Bias: A Randomized Experiment. *Science*. <https://doi.org/10.1126/science.1240466>. [Level 1; checked against abstract]
- Sacarny, Adam; Barnett, Michael L.; Le, Jackson; Tetkoski, Frank; Yokum, David; Agrawal, Shantanu (2018). Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults. *JAMA Psychiatry*. <https://doi.org/10.1001/jamapsychiatry.2018.1867>. [Level 1; checked against abstract]
- Sallis, Anna; Harper, Hugo; Sanders, Michael (2018). Effect of persuasive messages on National Health Service Organ Donor Registrations: a pragmatic quasi-randomised controlled trial with one million UK road taxpayers. *Trials*. <https://doi.org/10.1186/s13063-018-2855-5>. [Level 1; checked against fulltext]
- Shen, Lucas; Sood, Gaurav; Weitzel, Daniel (2026). Social Proof is in the Pudding: The (Non)-Impact of Social Proof on Software Downloads. *Journal of Online Trust and Safety*. <https://doi.org/10.54501/jots.v3i2.286>. [Level 1; checked against fulltext]
- Çat, Olgahan; Chang, Jiseon; Hlatky, Roman; Li, Huimin; Nielson, Daniel L (2024). Behavioral nudges in social media ads show limited ability to encourage COVID-19 vaccination across countries. *PNAS Nexus*. <https://doi.org/10.1093/pnasnexus/pgae189>. [Level 1; checked against fulltext]
- Çoker, Elif Naz; Pechey, Rachel; Jebb, Susan A. (2025). Testing the effect of a dynamic descriptive social norm message on meat-free food selection in worksite cafeterias: a randomized controlled trial. *BMC Medicine*. <https://doi.org/10.1186/s12916-025-04302-9>. [Level 1; checked against fulltext]
- Chapman, Cassandra M.; Spence, Jessica; Dixon, Lucas; Smith, Aimee E.; Hornsey, Matthew J. (2025). Social Norms and Charitable Giving: A Meta-Analytic Review. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/01461672251383996>. [Level 2; checked against fulltext]
- Mertens, Stephanie; Herberz, Mario; Hahnel, Ulf J. J.; Brosch, Tobias (2021). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.2107346118>. [Level 2; checked against fulltext]
- Papakonstantinou, Trisevgeni; Flecke, Sarah Lynn; Edmunds, C. E. R.; Cross, Rosina; Tran, Anh; Gold, Natalie (2025). A systematic review and meta-analysis of the effectiveness of social norms messaging approaches for improving health behaviours in developed countries. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-025-02275-6>. [Level 2; checked against fulltext]
- Tang, Mei Yee; Rhodes, Sarah; Powell, Rachael; McGowan, Laura; Howarth, Elizabeth; Brown, Benjamin; et al. (2021). How effective are social norms interventions in changing the clinical behaviours of healthcare workers? A systematic review and meta-analysis. *Implementation Science*. <https://doi.org/10.1186/s13012-020-01072-1>. [Level 2; checked against fulltext]

- Bohner, Gerd; Schlüter, Lena E. (2014). A Room with a Viewpoint Revisited: Descriptive Norms and Hotel Guests' Towel Reuse Behavior. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0104086>. [Level 3; checked against fulltext]

### Reviewer-suggested omissions for priming and ego depletion, listed but not extracted

- Priming: Chartier, Arnal, Arrow, Bloxson, Bonfiglio et al. 2020. Many Labs 5: Registered Replication of Albarracín et al. (2008), Experiment 5. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245920945963>
- Priming: Shariff, Willard, Andersen & Norenzayan. Religious Priming: A Meta-Analysis With a Focus on Prosociality. *Personality and Social Psychology Review* (Crossref issued 2015, print 2016). <https://doi.org/10.1177/1088868314568811>
- Priming: Shanks, Vadillo, Riedel, Clymo, Govind et al. 2015. Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000116>
- Priming: Chen, Latham, Piccolo & Itzchakov. An Enumerative Review and a Meta-Analysis of Primed Goal Effects on Organizational Behavior. *Applied Psychology* (Crossref issued 2020, print 2021). <https://doi.org/10.1111/apps.12239>
- Priming: Sanchez, Sundermeier, Gray & Calin-Jageman 2017. Direct replication of Gervais & Norenzayan (2012): No evidence that analytic thinking decreases religious belief. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0172636>
- Priming: Billingsley, Gomes & McCullough 2018. Implicit and explicit influences of religious cognition on Dictator Game transfers. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.170238>
- Priming: Vadillo, Hardwicke & Shanks 2016. Selection bias, vote counting, and money-priming effects: A comment on Rohrer, Pashler, and Harris (2015) and Vohs (2015). *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000157>
- Ego depletion / decision fatigue: Hagger, Wood, Stiff & Chatzisarantis 2010. Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*. <https://doi.org/10.1037/a0019486>
- Ego depletion / decision fatigue: Garrison, Finley & Schmeichel. Ego Depletion Reduces Attention Control: Evidence From Two High-Powered Preregistered Experiments. *Personality and Social Psychology Bulletin* (Crossref issued 2018, print 2019). <https://doi.org/10.1177/0146167218796473>
- Ego depletion / decision fatigue: Vadillo, Gold & Osman 2018. Searching for the bottom of the ego well: failure to uncover ego depletion in Many Labs 3. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.180390>
- Ego depletion / decision fatigue: Blázquez, Botella & Suero 2017. The Debate on the Ego-Depletion Effect: Evidence from Meta-Analysis with the p-Uniform Method. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2017.00197>

### Preprints excluded at full text (preprints were eligible only for multi-site replications)

- A Field Experiment Using Social Media Ads to Encourage COVID-19 Vaccination. <https://doi.org/10.21203/rs.3.rs-2319578/v1>. Reason: Preprint of a field experiment (criterion 5 allows preprints only for multi-site replications); superseded by the published version (Çat et al. 2024, *PNAS Nexus*, 10.1093/pnasnexus/pgae189), which is included.
- Social Proof is in the Pudding: The (Non)-Impact of Social Proof on Software Downloads. <https://doi.org/10.48550/arxiv.2603.07919>. Reason: Preprint duplicate of the published *Journal of Online Trust and Safety* article (included).
- Changing Real-World Behavior with Behavioral Requests and Social Norms Messages: A Meta-Analysis of Results from Field Experiments. <https://doi.org/10.21203/rs.3.rs-10244242/v1>. Reason: Preprint of a meta-analysis (criterion 5 allows preprints only for multi-site replications).
- Dynamic Norms Do Not Promote Vegetarian Food Preferences: Two Experiments and a Meta-Analysis. <https://doi.org/10.31234/osf.io/qfn6y>. Reason: Preprint (dynamic norms experiments plus small meta-analysis);

criterion 5.

- A Metastudy on Anchoring Effect Moderators. <https://doi.org/10.2139/ssrn.5936997>. Reason: Duplicate: SSRN version of the Schley & Weingarten metastudy preprint.
- A Metastudy on Anchoring Effect Moderators. [https://doi.org/10.31234/osf.io/tyq7f\\_v3](https://doi.org/10.31234/osf.io/tyq7f_v3). Reason: Not graded: preprint of a single-site preregistered online experiment (N = 1,967); preprints qualify only for multi-site replications (rule 4.5). Extracted as a non-grading row.
- Where are the context effects? A meta-analysis of the attraction, similarity, and compromise effects. [https://doi.org/10.31234/osf.io/v6y7e\\_v1](https://doi.org/10.31234/osf.io/v6y7e_v1). Reason: Not graded: preprint meta-analysis (rule 4.5). Extracted as non-grading context rows because it is the newest attraction-effect meta-analysis with bias adjustment.
- Hanging the Anchor Off a New Ship: A Meta-Analysis of Anchoring Effects. <https://doi.org/10.31234/osf.io/wf2tn>. Reason: Unpublished preprint meta-analysis (rule 4.5); not extracted.
- The Open Anchoring Quest Dataset: Anchored Estimates from 96 Studies on Anchoring Effects. <https://doi.org/10.5334/jopd.67>. Reason: Data paper (96 studies, N = 21,359); reports no pooled effect size with CI. Its effect-size meta-analysis is the unpublished Roeseler & Schuetz preprint.
- Does social psychology persist over half a century? (PsyArXiv preprint). <https://doi.org/10.31234/osf.io/w87yr>. Reason: Preprint of the published JPSP article 10.1037/pspa0000261; used only as the readable copy of that source, not counted separately.
- Does a \$1.20 Gift Pay for Itself? A Pre-Registered Randomized Field Experiment on Gifts in Direct Mail. <https://doi.org/10.2139/ssrn.7609039>. Reason: Preregistered field experiment with 1,950 US non-profit and educational organizations, but an SSRN preprint of a single-site study; rule 5 admits preprints only for multi-site replications. Abstract reports the gift lowered response (RR 0.732, 95% CI 0.603 to 0.889). Flagged for the lead because it is the only organizational-recipient test found and it points against the classic direction.
- Heterogeneous Effects of Loss Versus Gain Framing on Retirement Savings: Evidence from Two Randomized Field Experiments Using Causal Machine Learning. <https://doi.org/10.2139/ssrn.6812422>. Reason: Preprint (SSRN) of single-company randomized field experiments (N = 22,919 and 18,288); rule 5 admits preprints only for multi-site replications.
- Accelerating Psychological Science with Metastudies: A Demonstration Using the Risky-Choice Framing Effect. <https://doi.org/10.31234/osf.io/g76f9>. Reason: Preprint of DeKay et al. 2022 (Perspectives on Psychological Science, 10.1177/17456916221079611); a metastudy of many microstudies testing generalizability (N = 2,338), not a direct or multi-site replication or a meta-analysis of studies.
- Accelerating Psychological Science with Metastudies: A Demonstration Using the Risky-Choice Framing Effect. [https://doi.org/10.31234/osf.io/g76f9\\_v1](https://doi.org/10.31234/osf.io/g76f9_v1). Reason: Duplicate preprint version of 10.31234/osf.io/g76f9 (see that decision).
- Risky-Choice Framing Effects Persist When Option Descriptions are Matched and Complete: A Replication and Extension of DeKay and Dou (2024). [https://doi.org/10.31234/osf.io/nfwp3\\_v1](https://doi.org/10.31234/osf.io/nfwp3_v1). Reason: Preprint version of DeKay 2026 (Psychonomic Bulletin & Review); extracted from the journal version 10.3758/s13423-025-02771-w.
- Risky-Choice Framing Effects Persist When Option Descriptions are Matched and Complete: A Replication and Extension of DeKay and Dou (2024). [https://doi.org/10.31234/osf.io/nfwp3\\_v2](https://doi.org/10.31234/osf.io/nfwp3_v2). Reason: Preprint version of DeKay 2026; extracted from the journal version.
- Risky-Choice Framing Effects Persist When Option Descriptions are Matched and Complete: A Replication and Extension of DeKay and Dou (2024). [https://doi.org/10.31234/osf.io/nfwp3\\_v3](https://doi.org/10.31234/osf.io/nfwp3_v3). Reason: Preprint version of DeKay 2026; extracted from the journal version.
- Probabilities and the Risky-Choice Framing Effect: Evidence for an Inverted-U Pattern. <https://doi.org/10.31234/osf.io/vnq58>. Reason: Preprint of DeKay & Garge 2023 (Decision, 10.1037/dec0000214); an original study varying outcome probabilities, not a direct replication; preprint of a single-site study fails rule 5.

- Do Saving Nudges Work? A Pre-registered Field Experiment on Message Framing. <https://doi.org/10.2139/ssrn.7272436>. Reason: Excluded at title/abstract by the screener for  $N < 10,000$ , but a preregistered field experiment qualifies at any  $N$ ; re-checked: it is an SSRN preprint of a single-platform study, so rule 5 (preprints only for multi-site replications) excludes it.
- Effects of Positive-Negative Message Framing on Health-Related Attitudes, Intentions, and Behaviors: A Meta-Analysis. [https://doi.org/10.31234/osf.io/2gzum\\_v2](https://doi.org/10.31234/osf.io/2gzum_v2). Reason: Preprint of a meta-analysis (1,295 effect sizes, 430 studies); rule 5 admits preprints only for multi-site replications. Abstract reports very small framing effects ( $g = 0.04$  gain advantage for prevention;  $g = -0.08$  for detection).
- Searching for the bottom of the ego well (preprint version 1). [https://doi.org/10.31234/osf.io/ja2kb\\_v1](https://doi.org/10.31234/osf.io/ja2kb_v1). Reason: Duplicate preprint version; same reason as 10.31234/osf.io/ja2kb.
- The effect of ego depletion or mental fatigue on subsequent physical endurance performance: a meta-analysis (preprint). <https://doi.org/10.31234/osf.io/mr5pk>. Reason: Preprint of Giboin & Wolff 2019; context version; not retrievable (OSF blocked).
- Many Labs 5: Registered Replication of Albarracin et al. (2008), Experiment 7 (preprint). <https://doi.org/10.31234/osf.io/qzspr>. Reason: Duplicate preprint; the published version 10.1177/2515245920925750 is included (abstract only, not graded).
- Does willpower mindset really moderate the ego-depletion effect? A preregistered replication of Job, Dweck, and Walton (2010). <https://doi.org/10.1371/journal.pone.0287911>. Reason: Preregistered replication of Job, Dweck & Walton (2010) with  $N = 187$  (preprint abstract), below 500.
- Does mental fatigue impair physical performance? A replication study. <https://doi.org/10.1080/17461391.2020.1781265>. Reason: Context version (mental fatigue and physical performance); preregistered within-subject replication with 30 adults (preprint abstract), below the size rules.
- A Comprehensive Analysis of Choice Overload in Mutual Funds. <https://doi.org/10.2139/ssrn.4134300>. Reason: Working paper with no abstract; title indicates observational fund data, not an experiment; preprints of single studies are not eligible (rule 5).
- Too Much Choice? A Lab-in-the-Field Experiment on Choice Overload, Cognitive Constraints, and Consumer Welfare. <https://doi.org/10.2139/ssrn.7210669>. Reason: SSRN preprint of a single lab-in-the-field study,  $N = 284$  households (rules 3 and 5).