

Decision, Disagreement & Consensus Board

SAMPLE subject: three pricing offers for a design-review service

Prepared for: SAMPLE / EXEMPLU

Generated: 28 September 2026 · Confidence: Low

SAMPLE

Directional verdict: A leads, C polarizes, B is priced out.

Sample report with illustrative demo values. Not a real study, customer or market claim.

Not market validation. Not observed behavior. A structured synthetic read of where relevant personas may prefer, reject, misunderstand or disagree.

EXECUTIVE SUMMARY

- Variant A (€49 one-off) had the lowest average objection score in this SAMPLE run.
- Variant B triggered strong price resistance across budget-sensitive personas.
- Variant C split the panel: subscription fatigue versus low entry price.

DECISION

Bridge to action: test A's clarified wording with live traffic.

NEXT STEP

Move the strongest variant into a live SKPTST behavior test.

10 (SAMPLE)

Synthetic personas

3

Variants compared

0.41 (SAMPLE)

Consensus index

01 UNCERTAINTY FLAGS

- Small qualitative synthetic panel (SAMPLE n=10).
- Variant C wording differed in length.

02 PANEL CONSTRUCTION

Criterion	Spread
Buyer role	Founder, PM, freelancer
Budget sensitivity	Low / medium / high
Switching friction	Has agency vs none

03 SCORING MODEL

Comprehension is checked first as a disqualifying gate. Eligible options are summarized with median + range, a forced best-worst tally, segment cuts and polarization; disagreement is reported rather than averaged away.

04 USE CASE

Choosing which offer to put in front of real traffic first.

05 CONSENSUS BOARD

Directional median objection strength by variant (SAMPLE, 0–3).

	A · €49	B · €99	C · €19/mo
Price feels high	1	3	1
Unclear deliverable	2	1	2
Trust in reviewer	2	2	1
Subscription fatigue	0	0	3

06 CONFUSION WATCHLIST

- “Review” read as code review by some personas.
- “48h” unclear: start or delivery?

07 LANGUAGE TO TEST NEXT

- “10 prioritized design fixes in 48 hours”
- “One-off, no subscription”

METHODOLOGY

SAMPLE method: a 10-persona qualitative depth panel derived from the target audience and frozen as a quota matrix before the run. Options are shown blind, symmetrically and in randomized order. Personas answer independently, rejection reasons come first, comprehension is a hard gate (>2/10 material misreads disqualifies an option), aggregation uses median/range plus best-worst and segment/polarization cuts, and a diversity audit checks for mode collapse. Directional only, not statistical.

RESEARCH BEHIND THE METHOD

Current methodological context. These references are not evidence for this historical run.

MATCH PERSONA SOURCES TO THE DECISION

Lee, Dongryeol; Łajewska, Weronika; Perelli, Leonardo; Mansour, Saab (2026). Data-Driven Personas for Survey Simulation: Insights into Simulation Alignment Across Data-Access Regimes. arXiv preprint; accepted at REALM EMNLP 2026. Submitted: 5 October; revised: 6 October 2026.

Research finding: Across five survey datasets, eight persona sources and three models, unrelated-source personas rarely beat simple demographic context; correct group assignment and relevant history helped.

BEHAVIORAL HISTORY AND PERSONA CONTINUITY

Pourtaheri, Khashayar; Zareei, Ahmad (2026). Behavioral History Outperforms Descriptions of the Person for LLM Synthetic Personas. arXiv preprint. Submitted: 2 October 2026.

Research finding: In an 845-person two-wave study, past answers improved individual prediction more than demographic, personality and cognitive descriptions, while substantial individual signal remained missing.

EVALUATE OUTCOME TYPES AND INFERENCE COST SEPARATELY

Huang, Jin; Ferreras Garrucho, Diego; Xie, Yutong; Yuan, Walter M.; Mei, Qiaozhu; Lian, Chen; Hazell, Jonathon (2026). HouseholdBench: Evaluating Large Language Models as Predictors of Household Economic Behavior. arXiv preprint. Submitted: 6 October 2026.

Research finding: A 32-task benchmark compared 13 models with persistence and boosted-tree baselines. Fine-tuning plus prediction aggregation helped a small open-weight model on selected tasks; gains were not universal.

MAKE CALIBRATION ASSUMPTIONS INSPECTABLE

Gonzalez-Bonorino, Augusto; Biriukova, Kseniia; Capra, Monica (2026). Synthetic Cultural Agents from Aggregate Anchors. arXiv working paper. Submitted: 5 October 2026.

Research finding: Group-specific adapters retained declared aggregate preference signals, but transfer to new wording and agreement with human distributions were distinct and sometimes divergent.

EXPLICIT CONTEXT FOR SYNTHETIC PERSONAS

Argyle, L. P.; Busby, E. C.; Fulda, N.; Gubler, J. R.; Rytting, C.; Wingate, D. (2023). Out of One, Many: Using Language Models to Simulate Human Samples. Political Analysis, 31(3), 337–351. Online: 21 February 2023.

Research finding: Conditioned GPT-3 samples reproduced some patterns in studied US survey populations.

KEEP SIMULATION SEPARATE FROM POPULATION EVIDENCE

Bisbee, J.; Clinton, J. D.; Dorff, C.; Kenkel, B.; Larson, J. M. (2024). Synthetic Replacements for Human Survey Data? The Perils of Large Language Models. *Political Analysis*, 32(4), 401–416. Online: 17 May 2024.

Research finding: Synthetic survey responses showed compressed variation and sensitivity to prompts and time.

PANEL SIZE: 10 PROFILES, PLUS 2 CONTROL REPEATS

Guest, G.; Bunce, A.; Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59–82. February 2006.

Hennink, M.; Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. January 2022.

Research finding: In qualitative research with purposive samples, new themes stopped appearing within the first twelve interviews; a systematic review of 23 studies found saturation at 9–17 interviews for homogeneous populations and narrow objectives.

PROFILES CHOSEN BY QUOTA ON WHAT DRIVES THE DECISION

Guest, G.; Bunce, A.; Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59–82. February 2006.

Research finding: Purposive samples are the most common form of non-probabilistic sampling in qualitative research: participants are chosen for what they can reveal, not drawn at random.

EACH PROFILE SEES THE OPTIONS IN ITS OWN ORDER

Krosnick, J. A.; Alwin, D. F. (1987). An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly*, 51(2), 201–219. 1987.

Research finding: In a split-ballot survey experiment, the order in which answer choices were presented changed the answers (a primacy effect).

THE DECISION RULE IS FIXED BEFORE THE PANEL RUNS

Nosek, B. A.; Ebersole, C. R.; DeHaven, A. C.; Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. March 2018.

Research finding: Defining the questions and the analysis plan before seeing the outcomes separates tested predictions from explanations fitted after the fact.

FORCED CHOICE OF THE BEST AND THE WEAKEST OPTION

Finn, A.; Louviere, J. J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *Journal of Public Policy & Marketing*, 11(2), 12–25. September 1992.

Research finding: The study that introduced best–worst choice into marketing research: respondents name the most and the least important item in a set, which yields a relative order rather than separate ratings.

OBJECTIONS ARE ASKED BEFORE SCORES

Lord, C. G.; Lepper, M. R.; Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243. December 1984.

Research finding: In two experiments, asking people to consider the opposite possibility corrected biased judgment better than asking them to be fair.

SCORE WEIGHTS, WINNING MARGIN AND MODEL MIX

The method's own rule: The method sets the weights and the margin per type of decision and fixes them before each run: the criterion that matters most for that type of decision counts double, and a difference below the margin is treated as a tie.

CONTROL PRODUCT, DIFFERENCES INSTEAD OF LEVELS, AGREEMENT BETWEEN MODELS

The method's own rule: A product choice of the method, not validated on people: an everyday control product is asked the same way as the tested options, the panel is never shown a list of obstacles to rate, an option is placed against another only as a paired difference, and a difference on which the AI models point in different directions is not reported.

UNCERTAINTY NOTES

- Synthetic personas can share model biases.
- No real buyer behavior was observed.

EVIDENCE & SOURCES

Frozen persona quota matrix

Construction · 10 personas, independent blind runs (SAMPLE)

Variant copy

Input · Supplied by client (SAMPLE)