Neglect at Your Own Risk?

Evidence on Risk-Taking Prevalence and Motives from the Field

Saurabh Bhargava
New York University
sbhargav@gmail.com

Timothy Hyde
Oberlin College
thyde@oberlin.edu

This Version: August 2026

Abstract

We study risky choice in a field setting where employees choose among goal-reward contracts resembling financial lotteries and where we observe both choices and beliefs. We find risk aversion and choice heterogeneity far exceeding expected utility predictions and unexplained by prominent behavioral motives like overconfidence, nonlinear decision weights, and loss aversion. We propose and experimentally validate a heuristic explanation for risk taking involving contingency neglect during pairwise evaluation. The heuristic fits the field and lab data better than leading alternative models, uniquely predicts the belief distortions and framing effects we document, and offers a potential explanation for empirical insurance puzzles.

Acknowledgements: We extend a special thanks to Tim Houlihan and George Loewenstein for facilitating data access. We additionally thank Ned Augenblick, Linda Babcock, Karna Basu, Daniel Benjamin, Ben Bushong, Lynn Conell-Price, Stefano DellaVigna, Russell Golman, Kareem Haggag, Ben Handel, David Huffman, Alex Imas, Botond Koszegi, Yucheng Liang, Ted O'Donoghue, Ricardo Perez-Truglia, Alex Rees-Jones, Silvia Saccardo, Emmanuel Saez, Peter Schwardmann, Justin Sydnor, Lowell Taylor, Richard Thaler, Oleg Urminsky and seminar participants at UC Berkeley, Carnegie Mellon University, Hunter College, and New York University for constructive feedback. We also thank our partners at BI Worldwide for providing data and considerable program detail. We are especially appreciative of the generosity of Ray Harms, Jenn Kelby, Mark Hirschfeld, and Betsy Schneider. Finally, we thank Stephanie Rifai, Cassandra Taylor, and several research assistants for excellent project support. The authors had full editorial discretion and any errors are attributable to them.

1 INTRODUCTION

Why people take financial risks is central to economic theory and to the design of contracts and policies in insurance, employment, finance, and consumer protection. Under Expected Utility Theory (EU), risk aversion among fully informed, utility-maximizing decision-makers reflects diminishing marginal utility of wealth (von Neumann and Morgenstern, 1947). Yet observed choices often imply implausibly high utility curvature, exhibit more heterogeneity than standard models predict, and vary systematically across groups, including by gender (Rabin, 2000; Holt and Laury, 2002; Niederle, 2017).

Several alternative accounts of risk aversion have emerged, of which the most prominent emphasize biased beliefs, nonlinear decision weights, and loss aversion under reference-dependent preferences (Kahneman and Tversky, 1979; Loomes and Sugden, 1982; Gul, 1991; Prelec, 1998; Kőszegi and Rabin, 2006). Risk taking may also reflect heuristics, selective attention, or context-dependent evaluation rules less easily incorporated into utility-based frameworks (Kusev et al., 2017). Distinguishing among these accounts in the field is difficult: studies of insurance, investing, betting markets, and game shows document consequential risky behavior, but these settings typically involve complex decisions, limited observability of subjective beliefs, or limited generalizability (Barseghyan et al., 2018). We address these limitations by studying a setting that combines consequential field choices, directly elicited beliefs, and a tractable lottery structure, allowing us to distinguish among standard and nonstandard motives for risk taking.

Our data come from GoalQuest® (GQ), an employee goal-reward program designed by a consultancy to improve productivity at large North American firms. At the start of each one- to three-month program, employees privately select one of three performance goals, each associated with an all-or-nothing reward denominated in points. Menus typically feature roughly linear increases in goals and sharply convex increases in rewards, so the highest goal maximizes expected value under empirically reasonable attainment probabilities.

Three features make GQ valuable for studying risky choice. First, because performance outcomes are nested, goal choice is formally equivalent to choosing among nested lotteries over dollar-valued rewards: attaining Goal 3 implies attaining Goals 1 and 2, while failing to attain Goal 2 implies failing to attain Goal 3. In a preregistered validation experiment, participants choosing from a standard GQ menu behave similarly to those facing an economically equivalent, explicitly framed, and incentivized choice among nested lotteries. The setting therefore captures a general risky-choice problem rather than a peculiarity of workplace goal language. Second, through a research partnership, the consultancy temporarily added an onboarding module eliciting employees' perceived probability of attaining each goal immediately after choice, providing rare direct access to decision-relevant subjective beliefs. Third, the data have unusual scale and external relevance: we observe 20,133 employees across 34 field programs,

stakes ranging from $69 to $4,500, near-complete participation, and $9.4 million in rewards. Rewards are denominated in points redeemable for noncash goods in an online marketplace with a known point-dollar conversion rate; all dollar figures refer to the dollar-equivalent value of these points, and Section 3 addresses private reward valuation. We replicate the key descriptive patterns in an additional decision-only field sample of 15,345 employees and $8.2 million in rewards.

We document three empirical regularities in the field, using rational expectations as an initial benchmark. First, employees are substantially more conservative than the benchmark predicts: 49.2 percent choose a lower goal. Among these employees who attain at least Goal 1, the displayed-reward gap equals 45.3 percent of the reward associated with the benchmark goal when realized performance is held fixed. Second, employees exhibit substantially more heterogeneity in choice than the benchmark predicts: the Herfindahl–Hirschman Index of choice concentration is 0.350 in the data, compared with 0.747 under the benchmark. Third, conservatism differs sharply by gender. Women are about one-third more likely than men to choose conservatively, and this difference accounts for nearly all of the 22 percent female shortfall in realized rewards. Employees' elicited beliefs deepen the puzzle rather than resolve it: employees are, on average, optimistic — not pessimistic — about goal attainment. Taken together, these patterns — conservative choice relative to a standard benchmark, excess heterogeneity, and systematic gender differences, all despite overconfident beliefs — mirror prominent regularities documented elsewhere in the literature on risky choice.

To investigate the motives underlying these choices, we compare a broad set of models estimated on common data. We begin with the most prominent accounts of risk aversion — biased beliefs, nonlinear probability weighting, and reference-dependent loss aversion — alongside expected-value and expected-utility benchmarks. We then evaluate latent heterogeneity in risk preferences and several menu-based heuristics related to compromise, positional bias, and salience, along with reduced-form benchmarks for noisy or compressed probability judgments. The main specifications use CARA utility, and we replicate the principal results with CRRA utility in the appendix.

The results challenge each of the leading accounts. Utility curvature under rational expectations understates conservative choice and predicts excessive concentration in behavior. Incorporating subjective beliefs improves fit but does not explain conservatism: employees are overconfident on average, and this overconfidence is relatively greater for more ambitious goals— the opposite of the pattern required to rationalize the choices we observe. Nonlinear probability weighting adds little beyond subjective expected utility in the main specifications. Finally, a flexible loss-aversion specification improves in-sample fit but the implied degree of loss aversion is well above conventional empirical calibrations. These results suggest that the leading preference- and belief-based accounts leave an important decision process unexplained.

The shortcomings of these accounts motivate a different explanation. We posit that employees evaluate goals through pairwise comparisons between adjacent options. Such comparisons require conditional judgments: an employee comparing Goals 2 and 3, for example, should assess the probability of attaining Goal 3 conditional on attaining Goal 2. Research on conditional judgment suggests that people often neglect the conditioning event and anchor conditional probabilities toward unconditional base rates (Fox and Clemen, 2005; Sunstein and Zeckhauser, 2010; Martínez-Marquina, Niederle, and Vespa, 2019). Applied here, this error leads employees to anchor the conditional probability of attaining Goal 3 toward its lower unconditional probability, thereby understating the high goal's incremental value. We refer to this combination of adjacent pairwise evaluation and distorted conditional inference as pairwise contingency neglect (PCN). By compressing perceived value differences between adjacent goals, PCN naturally generates conservative and excessively heterogeneous choice.

We formalize this intuition in a parsimonious pairwise model of contingency neglect. PCN fits the field data substantially better than the benchmark preference-based alternatives by standard structural metrics, comes closest to matching the high prevalence of conservative choice and the excess dispersion of goal choices, and retains its advantage in held-out field programs. Focused comparisons indicate that this advantage is not explained by a reduced-form tendency to compress binary probability judgments toward 0.50, as can arise under cognitive uncertainty (Enke and Graeber, 2023), or by a generic penalty for choosing a more ambitious goal. Alternative pairwise accounts based on compromise, positional bias, and salience also perform materially worse (Simonson and Tversky, 1992; Christenfeld, 1995; Valenzuela and Raghubir, 2009; Bordalo, Gennaioli, and Shleifer, 2012, 2013). Estimated differences in contingency neglect also account for a meaningful share of the observed gender gap in conservative goals.

We conduct three complementary online experiments to probe the mechanism and address confounds inherent in the field data. Experiment A, a preregistered validation experiment, tests whether the central choice patterns persist when the GQ menu is presented as an explicit nested lottery, with and without real incentives, and whether presenting accurate pairwise conditional probabilities changes beliefs and choice. Experiment B moves beyond a single decision by placing participants in an incentive-compatible effort task and eliciting repeated choices and beliefs across six goal-reward menus that vary overall difficulty, the relative generosity of the high reward, and menu size. This within-person design tests whether stable preference-based explanations—including independently elicited loss aversion— and sorting heuristics based on perceived ability and competitiveness can account for behavior across menus, and whether PCN organizes the resulting menu-specific shifts in choice. Experiment C targets the inferential mechanism directly by comparing elicited conditional beliefs with participant-specific Bayesian benchmarks and varying whether the menu representation facilitates or impedes contingent inference. Together, Experiment A validates the lottery interpretation and corrective-information

prediction, Experiment B tests the competing models across repeated choices, and Experiment C directly examines the formation and behavioral relevance of distorted conditional beliefs.

The experimental evidence supports the process assumptions and descriptive predictions of PCN. Participants frequently report pairwise comparisons and behave consistently with pairwise evaluation of the menus. To examine whether the proposed inferential error extends beyond goal choice, Experiment C also elicits analogous conditional-probability judgments in a weather-forecasting domain. Across both goal-choice and weather judgments, participants systematically report a lower probability of the more extreme event conditional on the less extreme event than is implied by their own unconditional beliefs. To our knowledge, this is the first evidence that the conditional probability relevant to a binary pairwise comparison is systematically judged toward the corresponding marginal probability. Within the choice setting, the magnitude of this error predicts conservative goal selection, while presenting accurate conditional probabilities makes participants more likely to choose in accordance with expected-utility benchmarks. The experiments also replicate the field patterns of conservative and heterogeneous choice, and repeated choices in Experiment B vary across menus in the manner predicted by PCN rather than being organized by stable preference-based explanations.

Because many economically important choices involve nested contingent payoffs, PCN may apply beyond goal choice to insurance, portfolio allocation, contingent contracts, and tiered incentive schemes. We develop this extension for insurance, where coverage options ordered by cost sharing trade premiums paid in every state for benefits realized in progressively worse loss states. Evaluating an insurance upgrade therefore requires assessing both the probability of entering the relevant loss region and the severity of loss conditional on entering it. When ordinary experience directly represents the probability of entering that region but not its conditional upper tail, contingency neglect leads consumers to undervalue incremental protection and can produce underinsurance. When the loss region itself is rarely experienced, its probability must also be reconstructed; consumers evaluating coverage from within the loss branch may then fail to fully scale its benefits by the probability of reaching that branch. Under a common neglect process, this outer distortion dominates the opposing inner distortion and produces overinsurance. Section 7 formalizes how loss frequency and exposure history determine which probabilities must be reconstructed. The framework can thus help organize evidence of weak valuation of risk protection in prescription-drug plan choice (Abaluck and Gruber, 2011; Heiss et al., 2013) and unusually strong demand for low deductibles against modest, low-probability losses (Sydnor, 2010; Barseghyan et al., 2013).

We test these predictions in two hypothetical insurance experiments using different forms of equivalence. Experiment D, involving prescription-drug coverage, uses an informationally equivalent comparison: the presentations convey the same objective expenditure distribution but express the

comparison-relevant probability in marginal or conditional form. Directly supplying the accurate conditional probability raised selection of the benchmark-optimal plan by about 40 percent relative to baseline. This is the configuration the framework assigns to frequent, graded risks, where the outer probability is available and the conditional tail must be constructed. Experiment E, involving homeowners coverage, instead uses decision-equivalent frames: the presentations disclose different components of risk, but the same option remains benchmark-optimal over broad ranges of beliefs and risk preferences. Making the probability of no damage focal shifted choices toward that benchmark, whereas presenting only severity conditional on damage shifted choices toward more generous coverage; the share choosing more generous coverage increased by about 80 percent between the two presentations. Together, the experiments produce opposing directional shifts consistent with a common process, each matching the prediction assigned ex ante to the corresponding risk-and-information structure.

The paper contributes to four literatures. First, it advances research on the motives for financial risk taking in the field (Barseghyan et al., 2018) by showing that leading models based on biased beliefs, nonlinear probability weighting, and loss aversion leave important regularities unexplained when evaluated on the same field data. Second, it contributes to work on heuristics in menu-based financial decisions— including asset allocation (Benartzi and Thaler, 2007) and insurance choice (Ericson and Starc, 2012; Bhargava et al., 2017; Jaspersen et al., 2022) — by identifying a structurally tractable error in pairwise evaluation that is economically consequential and predictive across field and experimental settings. Third, it contributes to research on heterogeneity in risk taking, particularly gender differences (Niederle, 2017), by showing that differences in decision processes may contribute to behavioral differences traditionally attributed to preferences or beliefs. Fourth, by documenting systematic underestimation of pairwise conditional probabilities across choice and forecasting tasks, it connects evidence on failures of contingent reasoning — including neglect of correlation in school matching (Rees-Jones, Shorrer, and Tergiman, 2024) — to the literature on partition-dependent choice and inference (Fox and Rottenstreich, 2003; Benjamin, 2019; Ahn and Ergin, 2010). More broadly, the paper shows that risky choice in nested menus can depend materially on how decision makers locally represent conditional risk, not only on conventional belief and preference primitives.

2 BACKGROUND

2.1 Institutional Background

GoalQuest® (GQ) is an employee-rewards program conceived and administered by BI WORLDWIDE (BIW), a private global consulting firm specializing in incentive program design. Described as the world's only patented incentive-based sales program, GQ was designed to motivate employee productivity through self-selected performance goals tied to all-or-nothing non-monetary

rewards. As of 2026, BIW had administered over 1,000 GQ programs to over 1.4 million participants at firms primarily in the United States, Canada, and Europe since its inception in 2001. While marketed as a sales incentive program, our data indicate that the program has serviced employees engaged in a range of business functions (e.g., customer service, retention, investment) across a diversity of sectors including communication, health care, manufacturing, and financials.

2.2 Program and Goal-Reward Structure

GQ programs share a standardized structure: goal choice, goal attainment, and reward receipt.

Goal choice. During enrollment, employees are directed to an online portal where they privately select a goal from a menu of three personalized options (Goal 1, Goal 2, Goal 3), each associated with an all-or-nothing reward denominated in points (Appendix Figure A1). Goals were personalized to each employee's productivity: each menu was generated by applying a uniform rule to the employee's baseline performance, typically producing additively linear goals of the form $f(x_b)$, $f(x_b) + a$, $f(x_b) + 2a$, where $f(x_b)$ is a function of baseline productivity and a is a fixed increment.[1] Employees within a program were segregated into groups based on factors such as baseline performance, experience, or job level, with menus within each group personalized using the same rule. While goals scaled linearly, rewards increased in sharply convex increments—typically following a k, $3k$, $6k$ structure where k was approximately 1 percent of salary over the program duration. Combined with the all-or-nothing payoff structure, this convexity implies that the highest goal maximized expected value for the large majority of employees: using the early RE-EV benchmark defined in Section 3.1, Goal 3 was EV-maximizing for 85.8 percent of employees and Goal 2 for 8.5 percent. Employees were not given explicit encouragement to select any specific goal. BIW reports participation rates of 98 to 99 percent among eligible employees.

Goal attainment. During the subsequent 30- to 90-day performance period, employees work toward their selected goal, with programs providing access to intermediate performance data. In 2014, we asked BIW to implement an enhanced enrollment process to elicit employees' beliefs regarding goal attainment. Immediately after goal selection, employees were prompted to complete a brief optional survey asking them to estimate their perceived likelihood of attaining each goal: "On a scale from 0% (no chance) to 100% (absolute certainty), how likely is it that you will meet or exceed each of the following achievement levels?" (the response scale was indexed in 10-point increments). Employees were additionally asked about their binary gender, age, and tenure with the firm. While the survey was optional and rewards did not depend on completion, survey participation across our sample was 60 percent.

[1] Baseline performance was jointly determined by BIW and each firm based on factors such as data availability, employee tenure, and seasonal variation in productivity. For many programs, the baseline was calculated from employee performance over a recent period of similar duration to the program. New employees without historical performance were given a non-personalized menu.

Reward receipt. Employees who attain their goal exchange reward points for non-monetary rewards—including major electronics, vacations, and recreational items—in an online marketplace with a known conversion rate between dollars and points. Rewards were non-monetary, reflecting a belief that non-monetary incentives would be more motivating than cash rewards of similar value.

2.3 Data and Sample Construction

Primary Sample. We constructed the primary sample — 20,133 employees across 18 firms, 34 programs, and 232 groups — by applying screening restrictions to an original dataset ($n = 38{,}661$) reflecting all GQ programs administered between 2014 and 2018 in the US or Canada with enhanced enrollment, at least 100 fully participating employees, and electronically archived data.[2] We first excluded roughly 8% of employees with missing or inconsistent data or evidence of incomplete participation, yielding an expansive sample ($n = 35{,}478$). We then restricted to employees who completed enhanced enrollment with internally consistent beliefs to produce the primary sample.[3] In comparing the samples, employees completing enhanced enrollment were moderately more likely to select aggressive goals and modestly more likely to attain them, implying the conservatism and sub-optimal choice we subsequently document may, if anything, underestimate the actual degree of conservatism and sub-optimal choice in the broader employee population.[4] For robustness, we reproduce key analyses for the expansive sample in the Appendix. Collectively, employees in the primary (expansive) sample had the opportunity to earn $9.4 ($17.5) million in possible rewards. Appendix Table A1 overviews the primary sample and provides summary statistics by groups and employees.

Central Measures. Our analysis combines administrative data on employee goal choice and productivity with subjective beliefs elicited during enhanced enrollment. Appendix Table A2 summarizes employee choices, productivity, and goal attainment. Across programs, 44 percent of employees selected the highest goal, with a roughly even split across the remaining goals. We measure choice heterogeneity using the Herfindahl–Hirschman Index (HHI), which ranges from 1/n, representing maximal dispersion, to 1, representing complete concentration; the observed choice distribution yields an HHI of 0.350. Appendix Figure A2 presents choice shares across programs and groups. The elicited beliefs become the relevant probability inputs in our structural comparisons. Before turning to those comparisons, Section 3 constructs a separate rational-expectations benchmark to organize the descriptive field facts.

---

[2] Data for a small number of programs was not archived by BIW.

[3] An employee was tagged as having inconsistent beliefs if such beliefs implied a strictly greater likelihood of attaining a higher, relative to a lower, performance threshold. We excluded 2,215 employees, or 9.5% of enhanced enrollees, for this reason.

[4] We compared the expansive and primary sample across observable factors through regressions of the following form: $y_{i,l} = \alpha + \theta enhance_i + \pi_l + \varepsilon$, where y indicates an observable factor, enhance indicates completion of enhanced enrollment and π_l denotes group-level dummy variables. The most notable difference is that enhanced enrollees were 0.091 more likely to select Goal 3 (baseline choice share of 0.34) and 0.031 more likely to attain Goal 3 (baseline attainment of 0.28) than counterparts.

3 DESCRIPTIVE FACTS AND THEIR ROBUSTNESS

3.1 Descriptive Facts

We use expected value under rational expectations (RE-EV) as an initial benchmark for a series of descriptive facts. To construct the benchmark, for each employee in the primary sample, the attainment probability for each goal equals the mean attainment among all other employees in the same program group, a leave-one-employee-out construction. The benchmark's implications for choice classification are substantively unchanged when we instead use covariate-adjusted leave-one-employee-out predictions.[5] Table 1 reports the headline empirical implications, while Appendix Table A3 documents both constructions and additional sensitivity checks.

Conservative Choice. Under the primary RE-EV benchmark, Goal 3 maximizes expected value for 85.8 percent of employees, whereas 43.7 percent select it. Overall, 45.5 percent choose the RE-EV-optimal goal, 49.2 percent choose a more conservative goal, and 5.2 percent choose a more aggressive goal. Among employees classified as conservative who attained at least Goal 1, mean realized displayed reward is $165.81, compared with $303.24 at the RE-EV-optimal goal when realized performance is held fixed. The $137.43 gap is 45.3 percent of the counterfactual displayed reward.

Choice Dispersion. Observed choices are also substantially more dispersed than the RE-EV benchmark predicts. The observed HHI is 0.350, closer to the minimum of 0.333 under uniform choice than the 0.747 under the primary RE-EV prediction. Appendix Figure A3 shows that the difference is widespread across programs and groups. Thus, within the RE-EV benchmark, observable differences in estimated attainment probabilities generate far less heterogeneity than we observe. A sensitivity analysis indicates that explaining the observed heterogeneity solely as error in the RE benchmark would require large errors in employees' relative goal-attainment probabilities.[6]

Gender Differences in Conservative Choice. Women are 35 percent more likely than men to choose conservatively, a difference of 14.8 percentage points. This difference is economically meaningful: conditional on attaining Goal 1, women earn 22 percent less in realized rewards than men. A descriptive decomposition attributes 92 percent of this gap to gender differences in conservative choice.[7]

---

[5] Both constructions exclude the focal employee. The primary construction uses the mean attainment of the remaining employees in the focal employee's program group. The covariate-adjusted construction instead uses leave-one-employee-out regression predictions incorporating group effects and employee age and tenure. Relative to the primary construction, it changes the estimated conservative-choice share by approximately 0.2 percentage points and the share selecting the RE-EV-optimal goal by less than 0.5 percentage points. It modestly reduces predicted choice concentration, but the prediction remains substantially more concentrated than observed choice. Appendix Table A3 reports the complete comparison.

[6] For adjacent goals, EV rankings depend on the ratio of their attainment probabilities. Because attainment is nested, this ratio equals the probability of attaining the higher goal conditional on attaining the lower goal. Simulations that perturb these ratios reproduce the observed HHI only when mean absolute errors reach 28.9 to 33.3 percentage points.

[7] We estimate the share of the gender reward gap attributable to conservative choice by comparing the female coefficient in regressions of realized rewards with and without a control for conservative choice. We interpret the proportional decline in the coefficient after adding the control as the share of the gap explained by gender differences in conservative choice.

Overconfidence. Employees are optimistic, rather than pessimistic, about attaining each goal. Their reported attainment probabilities exceed the RE benchmarks by an average of 34.2 percentage points for Goal 1, 33.7 points for Goal 2, and 32.1 points for Goal 3. Absolute optimism alone, however, does not rule out a belief-based explanation for conservative choice, which would require relative underconfidence about more ambitious goals. We find the opposite. The ratio of mean subjective beliefs to mean RE probabilities rises from 1.78 for Goal 1 to 1.93 for Goal 2 and 2.09 for Goal 3, implying relative-overconfidence ratios of 1.09 for Goal 2 versus Goal 1, 1.08 for Goal 3 versus Goal 2, and 1.18 for Goal 3 versus Goal 1. Belief distortions therefore make ambitious goals relatively more, rather than less, attractive. Average belief gaps are similar for women and men.[8]

3.2 Alternative Explanations for the Descriptive Facts

Before interpreting these regularities through a model of choice under nested risk, we consider explanations specific to the field setting. GoalQuest is a workplace incentive program rather than a laboratory lottery: rewards are denominated in points, employees may value them differently, employees may possess private information about future productivity, and higher goals may require greater effort. We ask whether these features can account for the four regularities above — conservative choice, excess choice dispersion, gender differences in choice, and overconfident beliefs.

Reward Denomination and Valuation. Because GoalQuest rewards are denominated in points rather than dollars, conservative choice might reflect how employees translate points into private value. However, institutional details constrain this explanation. Each program provides a common-knowledge conversion between points and dollar-equivalent reward values. Holding effort costs aside, any proportional private discount to that conversion scales all three rewards equally and leaves their expected-value ranking unchanged. For example, valuing each point at 80 percent of its stated dollar equivalent converts a \$150–\$450–\$900 menu into a privately valued \$120–\$360–\$720 menu without changing the ranking. The same logic applies if proportional discounts vary across employees. Reward valuation can affect goal choice only if the marginal private value of points declines with reward size, thereby compressing the reward gradient and making higher goals relatively less attractive.

[8] For example, men and women had a near identical difference between perceived and actual Goal 3 attainment ($p = 0.73$).

Table 1.

Descriptive Characterization of Goal Choice

	All	Female	Male
N	20,133	9,335	10,794
Panel A. Characterization Overview			
Optimal Choice	0.46	0.40	0.50
Conservative Choice	0.49	0.57	0.42
Aggressive Choice	0.05	0.03	0.07
Observed HHI	0.35	0.34	0.37
Predicted HHI (RE-EV)	0.75	0.80	0.71
Panel B. Foregone Reward | Conservative Choice			
N	3,865	2,116	1,747
Mean Realized Reward	165.81	144.96	190.21
Mean Reward at RE-EV-Optimal Goal	303.24	266.53	346.86
Foregone Reward as Share of Optimal Reward	0.45	0.46	0.45
Foregone Reward as Share of Realized Reward	0.83	0.84	0.82
Panel C. Subjective Belief Gaps Relative to RE			
Mean Subjective Probability Minus RE Probability (% Points)			
Goal 1	34.18	32.85	35.34
Goal 2	33.72	32.41	34.85
Goal 3	32.08	31.10	32.93

Notes: This table reports descriptive facts for the primary field sample, using expected value under rational expectations (RE-EV) as the benchmark. RE probabilities are program-group leave-one-employee-out attainment rates; covariate-adjusted leave-one-out predictions are used for two singleton groups. The Female and Male columns use the imputed binary gender measure; four unclassified observations remain in All. Panel A reports observed and benchmark choice classifications and the Herfindahl-Hirschman Index (HHI). Panel B reports displayed reward differences among employees classified as conservative who attained at least Goal 1, holding realized performance fixed. Panel C reports the mean subjective attainment probability minus the corresponding RE probability in percentage points. Panel B is descriptive accounting, not a causal estimate of losses under counterfactual effort.

Two features of the setting weigh against this account. First, the redemption catalog is identical across goals within a program: higher goals provide more points for the same marketplace, not access to different rewards. Idiosyncratic tastes for particular goods therefore cannot generate goal-specific valuation without nonlinear valuation of points.[9] Second, declining marginal value predicts a monotonic increase in conservative choice with stakes. We find no such gradient. Dollar-equivalent potential rewards range from $69 to $4,500, but conservative choice is nearly identical in the top and bottom reward quartiles (54.1 versus 54.7 percent; clustered $p = 0.923$), and the program-group-level slope of conservative choice on reward value is small and statistically indistinguishable from zero. Structural estimates of choice bias are, if anything, smaller among higher-reward employees (Appendix Table A4). Thus, although nonlinear point valuation could in principle generate conservatism, the data provide no supporting gradient by financial stakes.

[9] Heterogeneous reward valuation could contribute to the gender gap if reward catalogs or tastes differed systematically by gender. The same institutional features constrain this account. Because catalogs are identical across goals and differ only in the number of points available, gender differences in tastes for particular rewards cannot generate gender differences in goal choice unless women's valuation of points is systematically more concave than men's—a specific and, to our knowledge, undocumented form of heterogeneity. We return to gender differences in the structural analysis, where we ask which decision primitives can account for the gap.

Private Information and Belief Measurement. An alternative possibility is that conservative choice relative to RE-EV might reflect information employees possess about their own ability and future performance conditions that is unavailable to the researcher. Employees' reported attainment probabilities should incorporate at least some of this private information. Yet using these reports to determine the expected-value-maximizing goal leaves the choice characterization virtually unchanged: 48.4 percent choose a lower goal, compared with 49.2 percent under RE-EV. As we show subsequently, allowing for utility curvature improves fit but still leaves substantial conservative choice unexplained. Results are also robust to shrinking subjective beliefs toward the RE benchmark.[10] As a complementary stress test, we construct attainment rates within program-group-by-chosen-goal cells, allowing for highly localized selection into goals. Conservative choice remains 43.1 percent, and this exercise rationalizes only 12.4 percent of choices classified as conservative under the primary benchmark (Appendix Table A3).

Endogenous Effort Costs. Finally, higher goals may require more effort, so employees could rationally choose a lower goal when the additional reward does not compensate for the additional cost. The timing of our belief elicitation, however, matters for evaluating this explanation. Employees report their attainment probabilities immediately after choosing a goal, so the natural interpretation is that these reports reflect the effort they intend to exert under the goal selected. A Goal 1 chooser, for example, may report her probabilities of attaining Goals 2 and 3 under the effort plan associated with Goal 1, rather than under the greater effort she would exert if she chose a higher goal. Her reported probabilities would then understate the attainability and expected value of those higher goals under the relevant counterfactual effort plans. Under this interpretation, our estimates understate the true prevalence of conservative choice: accounting for goal-specific effort would make higher goals more attractive and classify additional lower-goal choices as conservative. Appendix A.1 develops this intuition formally.

An alternative, though less natural, interpretation is that each reported probability already incorporates the optimal effort for its corresponding goal. The subjective expected-value benchmark would then capture the productivity benefit of greater effort but omit its cost, allowing effort costs to rationalize conservative choice in principle. Appendix A.2 evaluates this account by adding cumulative effort costs, expressed as a fraction of hourly wages, to the risk-neutral subjective expected-value benchmark. The calibration allows both an incremental cost of pursuing Goal 2 rather than Goal 1 and additional convexity in the cost of pursuing Goal 3. Appendix Table A5 shows that the model almost

[10] For each employee i and goal g, we construct adjusted beliefs $p\_ig(\lambda) = \lambda p\_ig^{sub} + (1 - \lambda)p\_ig^{RE}$, allowing λ to range from 0 to 1 in increments of 0.05. For each λ, we recompute the EV-maximizing goal and two expected-utility benchmarks with CARA parameters of 0.0003 and 0.005, then record the hit rate, conservative- and aggressive-choice shares, and predicted choice distribution. For all three benchmarks, the hit-rate-maximizing combination places full weight on subjective beliefs ($\lambda = 1$). Even then, the EV benchmark has a hit rate of 0.499 and classifies 0.484 of choices as conservative. The corresponding hit rates under EU with CARA parameters of 0.0003 and 0.005 are 0.501 and 0.532, with conservative-choice shares of 0.480 and 0.392. Shrinking noisy subjective beliefs toward rational expectations therefore does not rescue the standard benchmarks.

immediately produces corner predictions: low costs make Goal 3 optimal for most employees, while modestly higher costs make Goal 1 optimal for nearly everyone. In a one-month program, for example, a baseline effort cost of just 3 percent of hourly wages makes Goal 1 optimal for 91 to 93 percent of employees. The model can reproduce the observed interior mass of Goal 2 choices only implausibly, by assigning employees narrowly calibrated costs that covary precisely with their subjective beliefs.

Applied to employees' actual menus, this interpretation also predicts a pronounced correlation of choice with reward generosity relative to cumulative effort costs. We measure generosity as the Goal 3 reward per program workday, expressing rewards in the same per-workday units in which the calibration accrues wage-denominated effort costs. Actual choice is essentially invariant to this measure: in the bottom and top generosity quartiles, respectively, 29.1 and 29.2 percent of employees chose Goal 1, while 43.6 and 42.4 percent chose Goal 3 (mean goal choice: 2.15 versus 2.13; clustered $p = 0.896$). By contrast, with a baseline effort cost of only 1 percent of hourly wages, calibrated Goal 1 choice falls from 97–100 percent in the bottom generosity quartile to approximately 15 percent in the top quartile, while calibrated Goal 3 choice rises from 0–3 percent to 63–70 percent, depending on the convexity parameter. Taken together, the direction of bias under the chosen-effort interpretation, the corner predictions under the goal-specific-effort interpretation, and the absence of the predicted response to relative reward generosity make effort costs unlikely to account for the aggregate patterns. Effort undoubtedly affects productivity, but plausible effort costs cannot explain the combination of conservative choice and excess heterogeneity that motivates the structural analysis.

3.3 Is GoalQuest a Lottery-Like Choice? Evidence from Experiment A

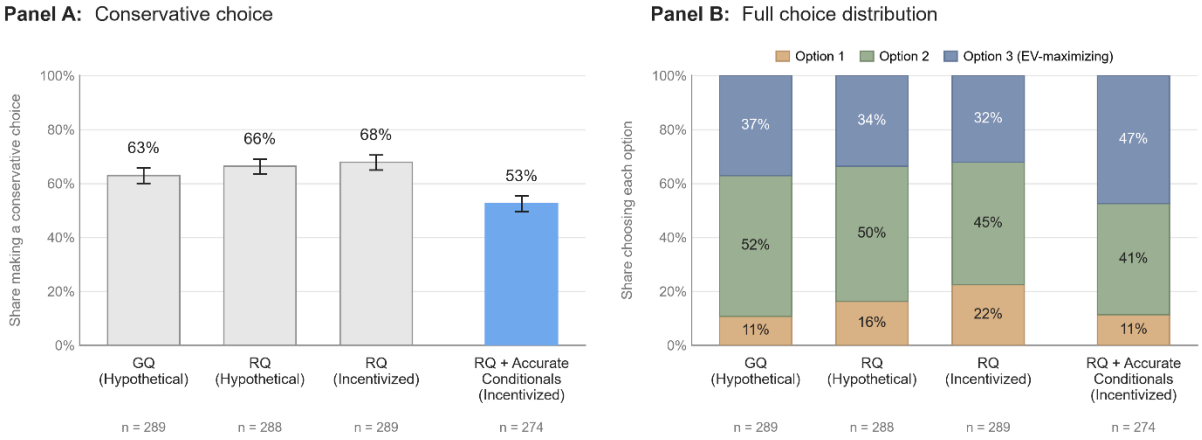
Experiment A stress-tests our interpretation of GoalQuest as a menu of nested risky prospects by comparing choices in a stylized GoalQuest setting with choices from an explicit menu of financial lotteries. The experiment addresses the field-specific explanations above, along with concerns about signaling, goal language, perceived managerial expectations, or confusion about the program. We represent GoalQuest as a set of options with increasing risk and reward that share a common source of uncertainty, such that attaining a more ambitious goal implies attaining each less ambitious goal, and test whether the central field patterns survive this stripped-down representation. The Experimental Appendix provides a schematic overview and key screenshots.

Experiment A was preregistered and administered in July 2026 to U.S. adults recruited through Prolific who were employed full- or part-time and were between 25 and 55 years old.[11] Of the 1,201 participants who began the study, 1,140 remained after the preregistered exclusions for potential bots and excessively fast completion. Participants were assigned with equal probability to one of four conditions.

[11] The preregistration is available at https://doi.org/10.1257/rct.19138-2.0.

In the GoalQuest condition (GQ), participants chose among goals of 105, 110, and 115 units, paired with rewards of $150, $450, and $900 and stated attainment probabilities of 83, 74, and 65 percent, calibrated to field averages. In the hypothetical RewardQuest condition (RQ), participants chose among three explicitly described nested lotteries with the same payoffs and probabilities. In the incentivized RewardQuest condition (RQ-I), the same lottery choice was consequential: conditional on a coin flip, participants could receive a bonus equal to one-half cent for each dollar of displayed reward earned. The fourth condition tests a mechanism developed later in the paper; we defer its results, together with the post-choice belief elicitations and decision-process measures, to that discussion. The first three conditions form a validation ladder. The GQ–RQ comparison asks whether replacing the GoalQuest frame with an explicit lottery changes choice, and the RQ–RQ-I comparison asks whether the hypothetical-lottery result survives real incentives. Option 3 has the highest EV in all three conditions: $585, compared with $333 for Option 2 and $124.50 for Option 1. We therefore classify Options 1 and 2 as EV-conservative.

Figure 1. Choice Across Representations, Incentives, and Probability Information (Experiment A)



Notes: Panel A plots conservative-choice shares; Panel B plots the full choice distribution. Conservative choice comprises Options 1 and 2; Option 3 maximizes expected value in all conditions. Error bars show ±1 SE. The sample applies the preregistered quality exclusions (N = 1,140).

Figure 1 shows that the experiment reproduces the central field regularities outside the workplace. The first three gray bars in Panel A indicate that conservative choice remains stable when the workplace frame is replaced by an explicit lottery and when choices are incentivized: 63.0 percent chose conservatively in GQ, 66.3 percent in RQ, and 67.8 percent in RQ-I. Panel B shows the full distribution of choices across Options 1–3 in each condition. Across the first three conditions, choice is highly heterogeneous: the HHI ranges from 0.360 to 0.422, compared with 1 under the deterministic EV benchmark. The experiment also reproduces the field gender difference: pooling the three conditions,

women were 16.0 percentage points more likely than men to choose conservatively.[12] These patterns were similar in magnitude to those in the closest field analogue.[13] Together, the results support the interpretation of GoalQuest as a nested risky-choice problem without establishing why participants depart from expected-value benchmarks. We return in Section 6 to the accurate-conditionals treatment shown in blue, along with the belief and process measures used to examine the proposed mechanism.

4 MODELS OF RISKY CHOICE

This section develops and compares alternative explanations for employees' goal choices. We begin by formalizing the RE-EV benchmark used to document the field patterns in Section 3.1. We then turn to a different question: taking employees' reported attainment probabilities as given, which model best explains their choices? Risk-neutral subjective expected value provides the starting point, and subjective expected utility provides the standard preference-based comparison. For each model, we describe how goals are evaluated and when the model predicts a choice below the risk-neutral choice implied by the same beliefs. We present the basic logic using two goals before extending each model to the three-goal menus observed in the field.

4.1 Rational-Expectations Benchmark (RE-EV)

Our framework describes the decision of a utility-maximizing employee choosing from a simplified menu of two productivity goals associated with all-or-nothing rewards. Because employee productivity varies across periods, each goal is naturally represented as a lottery, $G_n \in \{G_h, G_l\}$, yielding a reward x_n with probability s_n and no reward with probability $(1-s_n)$. The high goal has a strictly higher reward, $x_h > x_l$, and a lower likelihood of attainment, $s_h < s_l$. The productivity threshold associated with G_h exceeds that of G_l, and both are drawn from a common data-generating process, such that attainment of G_h implies attainment of G_l. We assume no discounting.

For this early benchmark, we assume employees hold rational expectations of goal attainment, denoted $\hat{s}_n^{re}$. If $u(\cdot)$ is a strictly increasing function mapping rewards to utility, normalized such that $u(0) = 0$, then a utility-maximizing employee selects a goal by solving: $\max_{n\in\{h,\,l\}}\ U(G_n) = \hat{s}_n^{re} \cdot$

[12] Pooling GQ, RQ, and RQ-I, 72.8 percent of women made an EV-conservative choice, compared with 56.8 percent of men, a difference of 16.0 percentage points (95 percent CI: 9.7–22.4; p < 0.001). The female–male difference was positive in every condition and did not differ significantly across conditions (interaction p = 0.31). Results are similar among participants who correctly answered both comprehension questions (N = 968): conservative-choice rates remain similar across conditions, and the pooled female–male difference is 17.6 percentage points (95 percent CI: 10.7–24.5; p < 0.001).

[13] The field analogue restricts the employee sample to menus with the same 1:3:6 reward gradient and to employees for whom Goal 3 maximizes expected value under rational-expectations probabilities. Conservative choice was somewhat less prevalent in the field analogue than across the three experimental conditions (59.9 versus 65.7 percent), whereas the female–male difference was similar (11.9 versus 16.0 percentage points). Choice was highly heterogeneous in both samples: the HHI was 0.341 in the field analogue and 0.387 in the experiment, compared with 1 under the deterministic EV benchmark.

$u(x_n)$. Under risk neutrality, $u(x) = x$. The employee therefore selects the low goal whenever $\hat{s}_l^{re} \cdot x_l > \hat{s}_h^{re} \cdot x_h$—that is, whenever the higher likelihood of the low goal compensates for its lower reward.

4.2 Standard Motives for Risk Aversion

A first explanation for conservative choice is diminishing marginal utility of rewards under standard expected utility. We incorporate risk aversion by adopting a utility function from the constant absolute risk aversion (CARA) family. If the parameter r captures attitudes toward risk ($r > 0$ implies risk aversion; $r = 0$ denotes risk neutrality; we restrict attention to $r \geq 0$), utility is given by:

$$u(x_n, r) = \begin{cases} \dfrac{1 - \exp(-r\, x_n)}{r}, & r > 0 \\ x_n, & r = 0 \end{cases}$$

This specification satisfies $u(0, r) = 0$ for all $r \geq 0$, preserving the expected utility representation from Section 4.1. The choice of a CARA function permits us to represent risk attitudes with a single parameter but implies the irrelevance of prior wealth for risk preferences. In Appendix A.3, we consider utility functions featuring constant relative risk aversion (CRRA) and show that this simplification does not affect goal characterization (Appendix Table A6). Following Alaoui and Penta (2025), we treat the resulting curvature estimate as a reduced-form parameter within the maintained utility representation rather than a self-interpreting psychological primitive: risk attitudes need not map one-for-one to the marginal value of rewards.

A utility-maximizing employee with concave utility selects the low goal whenever:

$$\hat{s}_l^{re} / \hat{s}_h^{re} > u(x_h, r) / u(x_l, r)$$

Because concavity compresses the utility of higher rewards relative to lower ones, $u(x_h, r)/u(x_l, r)$ decreases with r: as risk aversion grows, the utility advantage of the high goal's larger reward shrinks, making the probability advantage of the low goal more decisive. Conservative choice therefore becomes more likely as either the low goal's relative expected value or the employee's risk aversion increases. In the constrained specifications, we restrict r to [0,0.001]; at the upper bound, an employee would reject a 50/50 bet with infinite upside and a potential loss of $693. Because small-stakes choices can imply larger coefficients than conventional calibrations suggest, we treat this restriction as calibration discipline rather than as a definitive bound on risk aversion. To allow for heterogeneous preferences, we also estimate a three-parameter latent-class model.

4.3 Non-Standard Motives for Risk Aversion

Non-Standard Beliefs [$\hat{s}_n \neq \hat{s}_n^{re}$]. We next consider the possibility that conservative choice reflects systematic bias in employee beliefs of goal attainment. We model non-standard beliefs with a goal-specific multiplicative distortion, β_n, applied to the rational expectation, such that $\hat{s}_n = \beta_n \hat{s}_n^{re}$. Accordingly, $\beta_n > 1$ implies overconfidence while $\beta_n < 1$ implies under-confidence. A risk-averse, utility-maximizing employee with subjective beliefs selects the low goal whenever:

$$\hat{s}_l / \hat{s}_h > u(x_h, r)/u(x_l, r)$$

The decision rule implies that the likelihood of conservative choice is increasing in relative overconfidence of low versus high goal attainment. We assume subjective beliefs for remaining models.

Non-Standard Decision Weights [$\pi(\hat{s}_n) \neq \hat{s}_n$]. We proceed to consider whether non-linear decision weights might help explain conservative choice. We consider the inverse-S shaped probability weighting function proposed by Prelec (1998): $\pi(\hat{s}_n) = \exp(-(-\ln \hat{s}_n)^{\gamma})$, where γ governs the degree of curvature. Under non-linear decision weights, the low goal is selected whenever:

$$\pi(\hat{s}_l)/\pi(\hat{s}_h) > u(x_h, r)/u(x_l, r)$$

The decision rule implies that the likelihood of low-goal choice is increasing in the relative decision weight placed on low versus high goal attainment, π(ŝ_l)/π(ŝ_h), though the effect of γ on this ratio depends on the levels of ŝ_l and ŝ_h.

Loss Aversion [$v(y, \tau_n)$]. We next consider whether conservative goal choice may reflect prospective loss aversion in a model of gain-loss utility (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Although GoalQuest employees face prospective rather than realized losses, expectation-based models of gain-loss utility (e.g., Kőszegi and Rabin, 2006; Gul, 1991) and the interpretation of goals as reference points (Heath, Larrick, and Wu, 1999) suggest that loss aversion could favor conservative choice.

One challenge in using loss aversion to explain how goals influence risk taking is that the standard model assumes a single reference point. We therefore begin with a general value function defined over a generic outcome and a goal-specific reference point:

$$v(y, \tau_n) = \begin{cases} \eta\, m(y) + u^{+}(y - \tau_n), & \text{if } y \geq \tau_n \\ \eta\, m(y) + \lambda\, u^{-}(y - \tau_n), & \text{if } y < \tau_n, \end{cases}$$

where $m(\cdot)$ is an increasing consumption-utility function (nesting the CARA specification of Section 4.2 as a special case), u^{+} (concave) and u^{-} (convex) govern gain–loss utility in the gain and loss domains respectively, $\lambda > 1$ is the loss-aversion parameter, $\eta \geq 0$ scales the weight on consumption utility,

and $u^+(0) = 0$. To discipline these selections empirically, we compute the model's hit rate in the field across combinations of η and ten candidate reference points and retain the best-fitting specification (Appendix Table A7). This exercise selects $\eta = 1$ and a reference point equal to the reward from attaining the chosen goal, $\tau_n = x_n$. Intuitively, employees who select a goal adopt its reward as their expectation, so that attaining the goal leaves the employee at the reference point while failing to attain it places them in the loss domain. For the chosen-goal reference-point column in Table 2, the gain exponent is normalized to one and the fitted parameters are loss sensitivity λ, loss-side power curvature α, and logit noise σ; CARA r is not a parameter of that specification.

Under this specification, let $\Delta m_n \equiv m(x_n) - m(0)$ denote the consumption-utility gain from the reward and $\ell_n \equiv -u^-(-x_n) > 0$ the magnitude of gain–loss disutility from failing to attain goal n. The expected value of goal G_n is then $V(G_n) = m(0) + \hat{s}_n \, \Delta m_n - (1 - \hat{s}_n) \, \lambda \, \ell_n$. The employee selects the low goal whenever $V(G_l) > V(G_h)$, which, provided the denominator is positive, reduces to:

$$\lambda > \frac{\hat{s}_h \, \Delta m_h - \hat{s}_l \, \Delta m_l}{(1 - \hat{s}_h) \, \ell_h - (1 - \hat{s}_l) \, \ell_l},$$

Since high goals have lower success probabilities ($\hat{s}_h < \hat{s}_l$) and larger foregone rewards ($x_h > x_l$), and since u^- is increasing on the loss domain, it follows that $\ell_h > \ell_l$. Hence $(1 - \hat{s}_h)\ell_h > (1 - \hat{s}_l)\ell_l$, so the denominator is positive. As λ increases, the penalty from the high goal's larger expected loss grows, eventually tipping the balance in favor of the low goal. When the numerator is negative—that is, when the low goal already dominates on consumption utility—the condition is satisfied for all $\lambda \geq 1$, and loss aversion reinforces rather than drives conservative choice.

4.4 Pairwise Contingency Neglect

We propose a novel heuristic explanation for conservative goal choice which we call Pairwise Contingency Neglect (PCN). The model combines two well-documented behavioral tendencies: pairwise evaluation—the tendency to evaluate options through relative comparison—and failures of contingent reasoning, in which decision makers fail to condition correctly on hypothetical events (Esponda and Vespa, 2014; Martínez-Marquina, Niederle, and Vespa, 2019). The latter is closely related to partition dependence, whereby subjective probability judgments vary with how the state space is described (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Fox and Clemen, 2005).

PCN assumes that employees evaluate a menu of nested lotteries through a sequence of adjacent pairwise comparisons. The pairwise comparison, in turn, induces a partition of the state space that makes conditional reasoning central. When comparing two goals, the employee partially substitutes the unconditional probability of attaining the higher goal for the correct conditional probability of attaining it given that the lower goal has been reached. Because goals are nested, this substitution systematically

understates the conditional likelihood of high-goal attainment, generating overly conservative choice and excess choice heterogeneity relative to standard benchmarks.

Model Setup. Consider an employee comparing adjacent goals G_l and G_h, where $x_h > x_l$, $\hat{s}_h < \hat{s}_l$, and attainment of G_h implies attainment of G_l. The pairwise comparison partitions outcomes into three relevant states: S_L, output below G_l, in which the two goal choices are payoff-equivalent; S_M, output between G_l and G_h, in which the low goal pays x_l and the high goal pays zero; and S_H, output above G_h, in which both goals are attained but the high goal pays more. Let $\phi_{H|L+} = \hat{s}_h/\hat{s}_l$ denote the probability of attaining the high goal conditional on reaching the low one and let $\hat{\phi}_{H|L+}$ denote the employee's perceived conditional probability. The incremental utility gain from the high goal in state S_H is $\Delta u_h \equiv u(x_h, r) - u(x_l, r)$, while the incremental utility loss from the high goal in state S_M is $\Delta u_l \equiv u(x_l, r)$, where the second equality follows from normalizing $u(0, r) = 0$. The employee chooses the low goal whenever

$$\frac{\Delta u_l}{\Delta u_h} > \frac{\hat{\phi}_{H|L+}}{\left(1 - \hat{\phi}_{H|L+}\right)}$$

When $r = 0$, this rule reduces to the risk-neutral expected-value comparison for the adjacent pair. When $r > 0$, concavity compresses Δu_h relative to Δu_l, creating an additional force toward conservative choice.

PCN posits that the employee distorts the conditional probability towards the unconditional probability of the high goal: $\hat{\phi}_{H|L+} = (1 - \theta)\, \phi_{H|L+} + \theta\, \hat{s}_h$, where $\theta \in [0,1]$ indexes the severity of contingency neglect. When $\theta = 0$, the employee conditions correctly. When $\theta = 1$, the employee fully neglects the contingency. Since goals are nested, $0 < \hat{s}_h < \hat{s}_l < 1$, so $\hat{s}_h < \frac{\hat{s}_h}{\hat{s}_l}$. Any $\theta > 0$ therefore understates the probability of reaching the high goal conditional on reaching the low one, and correspondingly overstates the probability of the middle state S_M, where only the low goal pays. We estimate θ jointly with r, so the model nests standard pairwise expected utility ($\theta = 0, r > 0$), risk-neutral PCN ($r = 0, \theta > 0$), and the benchmark risk-neutral pairwise model ($r = \theta = 0$).

Comparative Statics. A higher θ shifts perceived probability mass from S_H, where the gain from upgrading is realized, toward S_M, where the cost of upgrading is realized. Rearranging the choice condition, the employee selects the low goal whenever

$$\hat{\phi}_{H|L+} < \frac{u(x_l, r)}{u(x_h, r)}.$$

The right-hand side is the threshold conditional probability required to justify upgrading. Under the convex reward structure typical of GoalQuest, approximately $(x, 3x, 6x)$, this threshold is $1/2$ for the G_2–G_3 comparison under risk neutrality, but only $1/3$ for the G_1–G_2 comparison. Moderate contingency neglect can therefore generate conservative choice at the top of the menu, while more severe neglect is

needed lower down. Utility curvature raises both thresholds, reinforcing this conservative force for any given θ. Because PCN compresses perceived value differences between adjacent goals, it also generates more choice heterogeneity than standard benchmarks predict, even absent heterogeneity in θ, because smaller perceived value differences amplify the role of idiosyncratic noise in the choice.

Worked Example. Consider an employee choosing between $G_l = (\$350, 0.85)$ and $G_h = (\$550, 0.60)$. The Bayesian conditional probability is $\phi_{H|L+} = 0.60/0.85 = 0.71$. Under risk neutrality, the upgrade threshold is: $\frac{u(x_l,0)}{u(x_h,0)} = \frac{350}{550} = 0.64$. A correctly conditioning employee therefore chooses the high goal. Under contingency neglect with $\theta = 0.75$, $\hat{\phi}_{H|L+} = 0.25 \times 0.71 + 0.75 \times 0.60 = 0.63$, so that the perceived conditional probability now falls below the upgrade threshold and the employee chooses the low goal.

Three-Goal Menus and Stopping Rule. With two goals, a pairwise comparison directly determines choice. With three goals, we must also specify how the two adjacent comparisons are combined. We assume that the employee first compares Goals 1 and 2. If Goal 1 is preferred, it is chosen; otherwise, Goal 2 is compared with Goal 3 and the preferred goal is chosen. We treat this ascending sequence as an implementation assumption rather than a separately identified mechanism. Participants' process reports support pairwise comparison, but they do not establish where comparisons begin, in what order they occur, or when deliberation stops. Consistent with this limited interpretation, Section 5.3 shows that the ascending sequential version of subjective expected value fits worse than its global counterpart. Thus, PCN's improved fit does not arise from the sequential structure alone.

4.5 Alternative Mechanisms Within the Pairwise Framework

PCN combines pairwise evaluation with a particular inferential error: the employee anchors the conditional probability of attaining the higher goal toward that goal's marginal attainment probability. To determine whether this specific error accounts for the model's performance, we compare PCN with alternatives that retain the same ascending pairwise choice rule but change either the perceived conditional probability or the value assigned to upgrading.

For an adjacent comparison between G_l and G_h, let $c_{lh} \equiv \varphi_{H|L+} = \frac{\hat{s}_h}{\hat{s}_l}$ denote the Bayesian conditional probability implied by the employee's reported attainment beliefs. Let $\Delta u_h \equiv u(x_h, r) - u(x_l, r)$ denote the utility gain from choosing the higher goal when both goals are attained, and let $\Delta u_l \equiv u(x_l, r) - u(0, r) = u(x_l, r)$ denote the utility loss from choosing the higher goal when only the lower goal is attained. For any perceived conditional probability q, define the net value of upgrading as $\Delta V_{lh}(q) \equiv \hat{s}_l[q, \Delta u_h, -, (1, -, q), \Delta u_l]$.

The undistorted pairwise value is therefore $\Delta V_{lh}{}^{0} \equiv \Delta V_{lh}(c_{lh})$and the employee upgrades whenever this value is positive. Under PCN, the perceived conditional probability is:

$$q_{lh}{}^{\text{PCN}} = (1, -, \theta)c_{lh} + \theta \hat{s}_h$$

where θ is between zero and one and governs the weight placed on the higher goal's marginal attainment probability; θ = 0 recovers the undistorted pairwise value.

Cognitive Uncertainty and Imprecision. One alternative is cognitive uncertainty - uncertainty about which judgment or action is correct - which can compress represented probabilities toward a 0.50 default (Enke and Graeber, 2023). We examine this reduced-form implication using a 0.50-compression model:

$$q_{lh}{}^{50} = (1, -, \kappa)c_{lh} + \frac{\kappa}{2}, \kappa \in [0, , , 1]$$

This model differs sharply from PCN. PCN's anchor is the employee- and goal-specific marginal probability ŝ_h, which is below the corresponding conditional probability in a nested menu. PCN therefore always reduces the perceived conditional probability. The 0.50-compression model instead uses a common anchor and can either lower or raise the conditional probability; when the Bayesian conditional is below 0.50, it predicts an upward distortion where PCN predicts a downward one.

We also estimate an adapted cognitive-imprecision model that allows employees to recover conditional probabilities noisily rather than systematically anchoring them on marginal beliefs:

$$q_{lh}{}^{\text{CI}} = \Lambda[\nu, \text{logit}, (c_{lh}), +, \tau, \varepsilon]$$

where ν governs systematic compression in log odds, τ governs imprecision, and ε is a standardized latent error over which choice probabilities are integrated. This is a flexible benchmark for noisy conditional reasoning rather than an exact structural implementation of any single theory of cognitive uncertainty.

Generic Upgrade Penalty. A separate possibility is that employees apply a common penalty to choosing a higher goal, independent of the probability structure of the menu. We represent this as a constant downward adjustment to each adjacent upgrade value:

$$\Delta V_{lh}{}^{P} = \Delta V_{lh}{}^{0} - \delta, \delta \geq 0$$

The employee upgrades whenever the adjusted value is positive. Unlike PCN, this specification leaves beliefs unchanged and subtracts the same amount from every adjacent upgrade. The models are distinguished by cross-sectional variation in PCN's belief distortion, which scales with the gap between the Bayesian conditional and the employee-specific marginal probability.

Encompassing Specification Checks. For the closest alternatives, we also estimate specifications that nest PCN and ask whether an additional component improves its fit. The first permits anchoring toward both the higher goal's marginal probability and 0.50: $q_{lh}{}^{\text{PCN+50}} = (1, -, \theta, -, \kappa)c_{lh} + \theta\hat{s}_h + \frac{\kappa}{2}$ where $\theta \geq 0, \kappa \geq 0,$ and $\theta + \kappa \leq 1$. The second adds a generic upgrade penalty to the PCN value:

$$\Delta V_{lh}{}^{\text{PCN+P}} = \Delta V_{lh}(q_{lh}{}^{\text{PCN}}) - \delta$$

We also estimate PCN jointly with the loss-aversion specification defined in Section 4.3. These are incremental-content tests rather than additional focal models. A positive κ or δ, or a loss-aversion estimate λ greater than one, would indicate that the corresponding mechanism explains variation beyond PCN; a boundary estimate would indicate that it does not. Appendix Table A8 reports the joint comparisons. The preceding comparisons focus on alternatives close to PCN's conditional-belief mechanism. We next consider three broader menu heuristics. These models retain Bayesian conditional beliefs and the ascending pairwise rule but alter the value assigned to adjacent upgrades.

Compromise Effect. Decision makers may favor interior options in ordered menus because they avoid the perceived extremity of the endpoints (Simonson and Tversky, 1992). We assign the middle option a position score of m_2 = 0 and the two extreme options scores of m_1 = m_3 = -0.25. The adjusted upgrade value is $\Delta V_{lh}{}^{C} = \Delta V_{lh}{}^{0} + \xi(m_h, -, m_l), \xi \geq 0$where ξ governs the strength of extremeness aversion. This adjustment adds 0.25ξ to the G_1-to-G_2 comparison and subtracts 0.25ξ from the G_2-to-G_3 comparison, pushing choice toward Goal 2 from both directions. The compromise effect can therefore explain excess Goal 2 choice, but it does not generate a general tendency to choose lower goals.

Positional Bias. Employees may also attach value directly to an option's ordinal position, apart from its reward and attainment probability (Christenfeld, 1995; Valenzuela and Raghubir, 2009). Because ordinal position enters through the two transitions in a three-goal menu, we estimate a separate shift for each adjacent comparison:

$$\Delta V_{12}{}^{B} = \Delta V_{12}{}^{0} + b_{12}$$

$$\Delta V_{23}{}^{B} = \Delta V_{23}{}^{0} + b_{23}$$

where b_12 and b_23 are freely estimated transition-specific positional shifts. Positive values favor advancing to the higher goal at the corresponding transition, while negative values favor stopping at the lower goal. Because these shifts do not depend on conditional probabilities, the model provides an identified reduced-form benchmark for menu-position effects distinct from contingency neglect.

Salience. Finally, choice may depend on the relative salience of rewards and attainment probabilities rather than on distorted conditional inference (Bordalo, Gennaioli, and Shleifer, 2012, 2013).

Let s-bar and x-bar denote the mean reported attainment probability and mean reward across the three goals. For each goal j, define probability and reward salience as $\omega_{s,j} = \frac{|\hat{s}_j, -, \bar{s}|}{\frac{(\hat{s}_j, +, \bar{s})}{2} + \zeta}$ and $\omega_{x,j} = \frac{|x_j, -, \bar{x}|}{\frac{(x_j, +, \bar{x})}{2} + \zeta}$ where ζ = 0.1 is a fixed regularization parameter. The salience-adjusted exponent is $a_j = 1 + \psi(\omega_{s,j}, -, \omega_{x,j}), \psi \geq 0$ and the resulting value of goal j is $V_j{}^S = \hat{s}_j{}^{a_j} u(x_j, r)$.

The salience-adjusted value of upgrading is therefore $\Delta V_{lh}{}^S = V_h{}^S - V_l{}^S$ and the employee upgrades whenever the adjusted value is positive. In estimation, a_j is bounded between 0.05 and 5 for numerical stability. This specification asks whether the relative distinctiveness of probabilities and rewards can explain adjacent choices without invoking errors in conditional inference. When ψ = 0, it reduces to standard pairwise expected utility. Appendix Table A9 compares these models using AIC and BIC, which account for their differing numbers of estimated parameters.

4.6 Estimation and Comparison.

Each model generates a latent value V_k^M for each goal $k \in \{1,2,3\}$ using employee-level inputs: subjective beliefs $(\hat{s}_1, \hat{s}_2, \hat{s}_3)$, personalized rewards (x_1, x_2, x_3), and model-specific parameters. The simultaneous-choice models—the risk-neutral benchmark, EU, Prelec, and loss aversion—compute V_k^M directly for all three goals. We map these values into choice probabilities using a multinomial logit:

$$P(k \mid M) = \frac{\exp(\sigma V_k^M)}{\sum_{j=1}^{3} \exp(\sigma V_j^M)},$$

where $\sigma > 0$ is a scale parameter governing choice sensitivity.

For the pairwise models—PCN and the alternatives in Section 4.5—choice probabilities follow from the sequential comparison rule. Let ΔV_{12} and ΔV_{23} denote the model-specific net values of upgrading from Goal 1 to Goal 2 and from Goal 2 to Goal 3, respectively. Applying logit noise yields:

$$P(G_1) = \Lambda(-\sigma \Delta V_{12}),$$

$$P(G_3) = [1 - P(G_1)]\ \Lambda(\sigma \Delta V_{23}),$$

$$P(G_2) = 1 - P(G_1) - P(G_3),$$

where $\Lambda(\cdot)$ is the logistic CDF. Goal 1 is chosen when the employee does not upgrade in the first comparison, Goal 3 when the employee upgrades in both, and Goal 2 otherwise. We estimate a separate scale parameter for each model so that fit comparisons are not driven by a common noise level.

All free parameters are estimated by maximum likelihood on the full primary sample. For the latent-class EU model, which allows for heterogeneous risk preferences within a three-type mixture (Section 4.2), we use expectation-maximization. Table 2 is the unrestricted main horse race and jointly estimates the ordinary curvature and behavioral parameters admitted by each model. Appendix Table A10 reports a separate coarse-grid, low-curvature calibration. Section 5.3 then uses a common constrained parameter space for a theory-disciplined PCN decomposition.

5 RESULTS FROM BENCHMARK HORSERACE

5.1 Full-Sample Comparisons

We compare the models on the primary sample, asking how well each explains three central features of behavior: conservative choice, the dispersion of choices across employees, and the shares choosing each goal. Table 2 reports the full-sample estimates. We organize the comparison around the classical account of risk aversion—utility curvature—and the leading behavioral alternatives—biased beliefs, probability weighting, and loss aversion—before turning to PCN.

The primary sample is representative of the broader field data. Appendix Table A11 shows that, under the rational-expectations benchmarks, the primary and expansive samples produce similar choice patterns and assign similar average probabilities to employees' observed choices. Employees without elicited beliefs behave much like those in the primary sample, so the comparisons that follow are not an artifact of the belief-elicitation subsample.

Table 2.
Structural Model Comparisons - Unrestricted Main Analysis

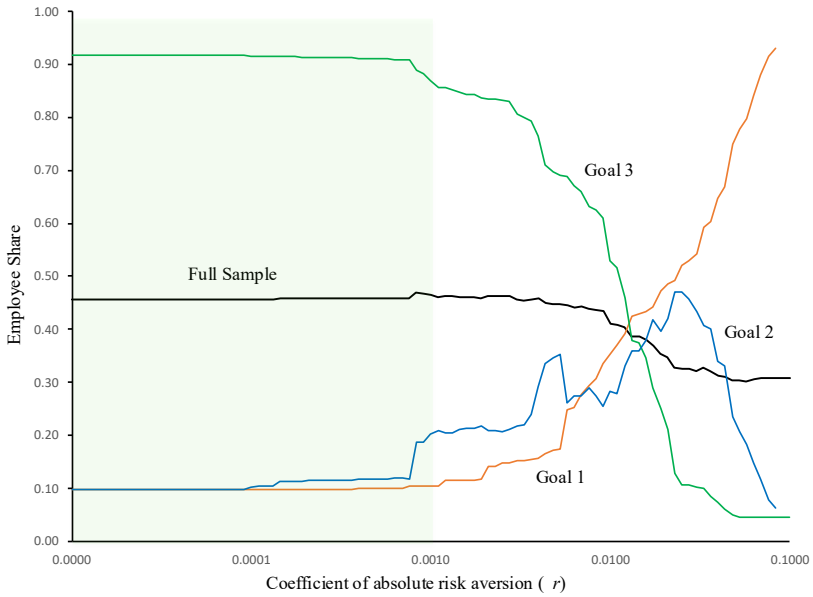
	Rational Expectations		Subjective Expectations					
	EV	EU	EV	EU	Latent-Class EU	RD-EU	LA	Pairwise CN
Model Fit Statistics								
Log Likelihood (LL)	-21812	-21515	-20935	-19805	-19674	-19701	-19560	-18662
AIC	43627	43035	41872	39613	39361	39408	39127	37330
BIC	43635	43051	41880	39629	39408	39431	39151	37354
LL Gain Relative to RE-EV	--	297	877	2008	2138	2112	2252	3150
LL Gain Relative to Subjective EU	-2008	-1711	-1130	--	130	104	244	1143
Hit Rate	0.46	0.45	0.50	0.53	0.53	0.54	0.59	0.58
Predicted Goal 3 Choice Share (Observed: 0.44)	0.86	0.72	0.87	0.75	0.75	0.74	0.56	0.51
Residual Conservative Choice Share	0.49	0.44	0.48	0.41	0.41	0.40	0.24	0.23
Share of RE-EV Gap Closed (1.00 = exact match)								
Conservative Choice	0.00	0.29	0.21	0.42	0.42	0.44	0.79	0.87
Herfindahl-Hirschman Index	0.00	0.45	-0.04	0.37	0.38	0.41	0.83	0.90
Key Parameters	--	rho = 0.0032	--	rho = 0.0038	rho = (0.044, 0.002, 0.005) class shares = (0.38, 0.15, 0.48)	alpha = 1.526 rho = 0.004	lambda = 4.87 alpha = 0.749	theta = 0.485 rho = 0.00293
Choice-Noise Scale (sigma)	419.28	67.10	315.73	53.18	25.24	53.37	180.73	31.64

Notes: This table compares structural models in the primary field sample. Rational-expectations (RE) probabilities are program-group leave-one-employee-out attainment rates, with covariate-adjusted leave-one-out predictions for two singleton groups; subjective-expectations models use elicited beliefs. EV denotes expected value; EU denotes CARA expected utility; Latent-Class EU is a three-type mixture; RD-EU uses Prelec probability weighting; LA denotes loss aversion; and Pairwise CN denotes pairwise contingency neglect. Reported statistics are log likelihood, AIC, BIC, deterministic hit rate, predicted Goal 3 share, residual conservative-choice share, and closure of the RE-EV-to-data gaps in conservative choice and the Herfindahl-Hirschman Index (HHI). Residual conservative choice means choosing below the model's deterministic prediction; gap closure equals 1 at an exact match, 0 at RE-EV, is negative when fit moves away from the data, and can exceed 1 with overshooting. Pairwise CN jointly estimates theta, CARA rho, and logit noise sigma. Chosen-goal-reference LA instead estimates lambda, loss-side power alpha, and sigma, normalizing the gain exponent to one. The reported profile sets are one-dimensional grid profiles; the estimated theta-rho correlation is -0.546. Relative fit applies only to the implemented model set.

Utility curvature alone cannot account for the observed pattern. Standard EV and EU models under rational expectations predict Goal 3 shares of 0.86 and 0.72, compared with an observed share of 0.44, and leave a large share of employees classified as choosing conservatively. Diminishing marginal utility helps—in Table 2, EU closes 29 percent of the conservative-choice gap and 45 percent of the HHI gap—but these gains fall to 17 and 28 percent under a low-curvature calibration (Appendix Table A10), and the model still predicts far too much Goal 3 choice. Figure 2 shows why: as risk aversion increases, predicted Goal 1 choice rises while predicted Goal 3 choice falls, and these offsetting shifts leave the share of observed choices the model rationalizes nearly unchanged.

Subjective beliefs substantially improve fit but do not explain conservative choice. Under expected value, replacing rational-expectations probabilities with employees' reported probabilities raises the log likelihood by 877 points, isolating the contribution of beliefs while holding utility fixed. Yet the predicted Goal 3 share rises from 0.86 to 0.87: the improvement comes from matching other features of the choice distribution, not from explaining lower-goal choice. The direction of the belief errors explains why—employees are relatively optimistic about attaining higher goals, whereas a beliefs-based account of conservatism requires relative pessimism. Allowing utility curvature helps: unconstrained subjective EU closes 42 percent of the conservative-choice gap and 37 percent of the HHI gap. Yet it predicts a Goal 3 share of 0.75 against the observed 0.44 and requires $\rho = 0.0038$, nearly four times the upper bound adopted in Section 4.2. Restricting ρ to that range raises the predicted Goal 3 share to 0.84 and reduces the corresponding gap closures to 26 and 7 percent (Appendix Table A10).

Figure 2.
Optimal Choice under Expected Utility with Rational Expectations



Notes: This figure plots the share of employees for whom each goal, and any goal overall, is optimal under the expected utility benchmark as the CARA coefficient, r, varies on a logarithmic scale. The shaded region denotes the plausible range $r \in [0, 0.001]$.

The remaining behavioral accounts fare little better. A three-type expected-utility model and nonlinear probability weighting yield only modest gains over subjective expected utility. Loss aversion performs best among them, but the fitted model requires a loss-aversion coefficient of 4.87—well above conventional calibrations—and flexible curvature on the loss side.

PCN is the best-fitting model in Table 2 by a substantial margin. It gains 1,143 log-likelihood points over unconstrained subjective expected utility and nearly 900 over loss aversion, and it has the lowest AIC and BIC among the models considered. PCN predicts a Goal 3 share of 0.51 against the observed 0.44 and closes 87 percent of the conservative-choice gap and 90 percent of the HHI gap. The estimated PCN parameter of 0.485 implies that employees weight the marginal probability of attaining the higher goal about as heavily as the correct conditional probability. Adding loss aversion to PCN yields limited, specification-dependent gains: the joint model gains about 26 log-likelihood points when curvature is fixed, but when curvature is re-estimated, the loss-aversion coefficient equals one and improves neither full-sample fit nor held-out fit in any of the 34 leave-one-program-out comparisons (Appendix Table A8).

These comparisons are stable across subgroups. In Appendix Table A4, all models fit better among employees facing larger rewards and among more experienced employees, consistent with choices becoming more systematic as stakes and familiarity increase. PCN is the best-fitting model in every subgroup, and estimated contingency neglect is smaller at higher rewards and with greater experience. Loss-aversion estimates, by contrast, vary sharply across subgroups and sometimes fall below one, making them difficult to interpret as stable preference parameters.

5.2 Out-of-Sample Validation

The full-sample results show that PCN fits the observed choices well; Table 3 asks whether it generalizes to new programs. For each of the 34 programs, we estimate the models using employees in the other 33 programs and predict choices in the omitted program, so every reported prediction is made for a program not used in estimation. PCN predicts held-out choices substantially better than subjective EU and RE-EV. Its total held-out log score is −18,808, compared with −19,927 for subjective EU and −21,865 for RE-EV, where a less negative score indicates better predictions. It correctly predicts 57.5 percent of held-out choices, compared with 52.3 percent for subjective EU and 45.6 percent for RE-EV, and it outperforms subjective EU in 26 of the 34 programs, with an average program-level log-score advantage of 32.91 points (95% CI: 14.22 to 51.61; $p = 0.0011$). The advantage is not confined to a few large programs. The held-out predictions also reproduce the central features of the choice distribution. PCN closes 89.6 percent of the RE-EV gap in conservative choice and 90.9 percent of the HHI gap, compared with 42.1 and 38.2 percent for subjective EU.

Table 3.

Central Held-Out Model Comparison - Pairwise CN versus Subjective Expected Utility

Full-Sample Fit and Leave-One-Program-Out Prediction	RE-EV	Subjective EU	Pairwise CN
Full-Sample Log Likelihood	-21,812.43	-19,804.73	-18,662.09
Leave-One-Program-Out Log Score	-21,865.04	-19,927.34	-18,808.28
LOPO Log-Score Gain Relative to RE-EV	--	1,937.70	3,056.76
LOPO Log-Score Gain Relative to Subjective EU	-1,937.70	--	1,119.06
Mean Held-Out Log Score per Choice	-1.086030	-0.989785	-0.934201
Held-Out Hit Rate	0.456	0.523	0.575
Mean Probability Assigned to Observed Choice	0.344	0.409	0.449
Predicted Goal 3 Choice Share (Observed: 0.437)	0.862	0.749	0.501
Predicted Conservative Choice Share (Observed: 0.494)	0.000	0.208	0.443
Residual Conservative Choice Share	0.494	0.415	0.225
Predicted HHI (Observed: 0.350)	0.752	0.598	0.386
Share of RE-EV Conservative-Choice Gap Closed	0.000	0.421	0.896
Share of RE-EV HHI Gap Closed	0.000	0.382	0.909
Paired Program p-value: PCN vs Subjective EU	--	--	0.0011
Paired Program Mean Log-Score Gain: PCN vs Subjective EU (95% CI)			32.91 [14.22, 51.61]
Key Parameters	sigma = 419.279	rho = 0.003754 sigma = 53.213	theta = 0.4845 rho = 0.002931 sigma = 31.638

Notes: This table reports leave-one-program-out predictive comparisons for 20,133 employee choices in 34 programs. Each fold re-estimates the model on 33 programs and scores the omitted program. Subjective EU weights CARA utility by elicited attainment probabilities and uses multinomial-logit choice noise. Pairwise CN uses the same beliefs, Bayesian adjacent-goal conditionals, CARA utility, and an ascending sequential-logit choice rule, re-estimating theta, rho, and sigma in every fold. The confidence interval and two-sided p-value use paired program-level held-out log-score differences. Residual conservative choice is the share choosing below the model's deterministic prediction. Gap closure compares pooled held-out predictions with the observed pooled moment, setting RE-EV to 0 and an exact match to 1. The predictive comparison is limited to the two models re-estimated in the grouped folds.

5.3 Decomposing the Contribution of PCN

Table 4 decomposes the PCN model, building it in stages from subjective expected value to isolate what each component contributes. We first introduce the ascending pairwise choice rule, add modest utility curvature, and finally add contingency neglect. After the first step, all models use the same ascending rule and restrict curvature to plausible values spanning the interval from zero to 0.001, so that changes in fit can be attributed to the component added at each stage.

The pairwise rule alone worsens fit, lowering the log likelihood by 322 points relative to global subjective expected value. Constrained curvature more than offsets this decline, improving the log likelihood by 1,258 points, and contingency neglect adds a further 819 points. The final step also moves the predicted Goal 3 share from 0.526 to 0.437—essentially matching the observed 0.44—and improves the model's account of both conservative choice and dispersion. PCN's advantage therefore comes from the distortion of conditional beliefs, not from the sequential choice structure itself, which alone worsens fit. The estimated PCN weight in this exercise is 0.812, exceeding the unconstrained estimate of 0.485 because the calibration restricts curvature, leaving the neglect parameter to absorb more of the fit.

Table 4.

Pairwise Contingency Neglect - Theory-Disciplined Mechanism Decomposition (Constrained Curvature)

	Subjective EV (global)	Subjective EV (sequential)	Subjective EU (sequential)	Pairwise CN
Model Fit Statistics				
Log Likelihood (LL)	-20935	-21257	-19999	-19180
ΔLL Relative to Global Subjective EV	--	-322	936	1755
ΔLL Relative to Sequential Subjective EV	--	--	1258	2077
ΔLL Relative to Sequential Subjective EU	--	--	--	819
Hit Rate	0.50	0.45	0.52	0.56
Predicted Goal 3 Choice Share (Observed: 0.44)	0.87	0.37	0.53	0.44
Residual Conservative Choice Share	0.48	0.17	0.24	0.18
Share of RE-EV Gap Closed (1.00, exact match)				
Conservative Choice	0.21	1.18	0.88	1.03
Herfindahl-Hirschman Index	-0.02	0.63	0.76	0.88
Behavioral / curvature parameters			rho = 0.001	theta = 0.812
Additional parameter				rho = 0.001
Choice-noise scale (sigma)	315.7	159.3	77.9	53.5

Notes: This table decomposes the pairwise-contingency-neglect model in the primary sample of 20,133 employee choices. The first column is global subjective EV with multinomial-logit choice noise; the remaining columns share an ascending sequential-logit architecture. Sequential subjective EV adds the implementation rule, sequential subjective EU then adds CARA curvature, and Pairwise CN adds the contingency-neglect parameter theta. All sequential curvature specifications impose rho in [0, 0.001], and the upper bound binds. Relative to the preceding column, sequential stopping changes log likelihood by -322.26, constrained curvature adds 1,258.22 points, and Pairwise CN adds 819.00 points. The final increment isolates PCN within the disciplined sequential architecture; it does not imply that the stopping rule is the dominant mechanism. The constrained estimate theta = 0.812 is conditional on the binding curvature cap and is not an unrestricted field estimate.

5.4 Alternative Pairwise Accounts

The results so far show that PCN outperforms the classical and leading behavioral accounts of risk aversion. A remaining question is whether its advantage reflects the specific mechanism we propose—anchoring conditional probabilities on marginal probabilities—or other departures from expected utility that share the same pairwise structure. We consider five alternatives: compression of probability judgments toward 0.50, as can arise under cognitive uncertainty (Enke and Graeber, 2023); a generic penalty on choosing a more ambitious goal; and pairwise versions of compromise effects, positional bias, and salience. Table 5 and Appendix Table A9 estimate these alternatives separately; Appendix Table A8 tests whether the closest alternatives—loss aversion, a common 0.50 anchor, and generic upgrade penalties—add explanatory power when estimated jointly with PCN.

Table 5.

Focused Field Model Comparisons and Falsification Checks - Common Low-Curvature Calibration

	Sequential Models (rho = 0.001)							MNL Model
	Pairwise EU	50% Compression	Cognitive Imprecision	Generic Upgrade	Pairwise CN	PCN + Upgrade Penalty	PCN + Loss Aversion (fixed	Canonical CPT (MNL)
Model Fit Statistics								
Log Likelihood (LL)	-19999	-19962	-19872	-19999	-19180	-19180	-19154	-19270
AIC	40001	39928	39751	40003	38365	38367	38315	38549
BIC	40008	39944	39775	40018	38380	38390	38338	38580
LL Gain Relative to Pairwise EU	--	37	127	0	819	819	845	729
LL Gain Relative to Pairwise CN	-819	-782	-692	-819	--	0	26	-90
Hit Rate	0.52	0.52	0.53	0.52	0.56	0.56	0.55	0.60
Predicted Goal 3 Choice Share (Observed: 0.44)	0.53	0.52	0.54	0.53	0.44	0.44	0.43	0.53
Residual Conservative Choice Share	0.24	0.24	0.24	0.24	0.18	0.18	0.18	0.23
Share of RE-EV Gap Closed (1.00 = exact match)								
Conservative Choice	0.43	0.43	0.42	0.43	0.51	0.51	0.52	0.42
Herfindahl-Hirschman Index	0.75	0.78	0.75	0.75	0.88	0.88	0.88	0.88
Key Parameters	rho = 0.001	kappa = 0.189	precision = 0.999 tau = 1.369	delta = 0.00	theta = 0.812	theta = 0.812 delta = 0 (95% UB: 0.09)	theta = 0.665 lambda = 1.224	lambda = 1.002; alpha = 0.610; gamma = 1.724
Choice-Noise Scale (sigma)	77.93	61.58	68.38	77.93	53.52	53.52	54.46	14.92

Notes: This table reports focused comparison and falsification tests in the primary sample of 20,133 employee choices, using elicited subjective beliefs. Except for canonical CPT, all columns use an ascending sequential-logit architecture and impose CARA rho = 0.001 so differences isolate the displayed distortion. Pairwise EU is the architecture-matched expected-utility comparator; 0.50 Compression shrinks conditional probabilities toward 0.50; Generic Upgrade Penalty adds a constant or proportional cost of moving to a higher option; and the joint columns add each candidate mechanism to Pairwise CN. Canonical CPT uses multinomial logit and Prelec probability weighting. Added anchor and penalty parameters are restricted to be nonnegative, and boundary estimates reproduce the nested comparator. Delta log likelihood is measured relative to the relevant nested model, while lower BIC indicates preferred penalized fit. The common-MNL CPT comparison uses paired program-level log-score differences. These are targeted falsification tests rather than an exhaustive set of behavioral models.

None of these alternatives approaches PCN's fit. Compression toward 0.50 improves the log likelihood by only 37 points relative to pairwise subjective expected utility and remains 782 points behind PCN; when the marginal-probability and 0.50 anchors are included together, the estimated weight on 0.50 is zero whether curvature is fixed or freely estimated. The generic upgrade penalty is likewise estimated at zero both on its own and when added to PCN. A proportional version estimates a small 1.65 percent penalty and gains only 1.96 log-likelihood points, insufficient to offset its added complexity: BIC worsens by 6.00. Compromise effects and positional bias improve modestly on pairwise subjective expected utility but remain far behind PCN and continue to predict too little conservative choice, and the salience parameter is estimated at essentially zero. PCN's advantage is therefore not a generic consequence of probability compression, reluctance to choose higher goals, middle-option attraction, ordinal position, or attribute salience. The experiments in Section 6 provide additional tests of PCN against these alternative motives, examining whether choices shift across menus in the pattern PCN predicts rather than following stable preferences or position effects.

5.5 Robustness to Effort-Dependent Beliefs and Costs

Section 3 showed that effort costs do not explain the choice patterns whether the post-choice belief reports reflect the effort planned under the chosen goal or the optimal effort the employee would exert under each respective goal. Appendix Table A12 examines a more flexible version of the effort account that allows attainment beliefs and effort costs to adjust jointly across goals. For each employee,

we treat the reported belief vector as reflecting the effort associated with the chosen goal and construct counterfactual beliefs and costs for each unchosen goal, allowing the employee to exert more or less effort than under the observed choice. Across 65 combinations, we vary both the effect of these effort changes on attainment probabilities and the level and convexity of effort costs, estimating each model on the training sample under every combination and evaluating its predictions in the holdout sample. These adjustments do not improve out-of-sample prediction. The highest holdout hit rate anywhere in the grid is no better than the no-effort benchmark for any model except rational-expectations EU, where the improvement is less than one-tenth of a percentage point. More importantly, selecting an effort specification on training-sample fit never improves the holdout mean log score: it leaves the rational-expectations and loss-aversion models unchanged and worsens the models using subjective beliefs. The effort adjustments fail for the same reason as the simpler calibrations. Greater effort raises the probability of attaining a higher goal, while the associated cost makes that goal less attractive. These forces either offset one another or push predicted choices toward Goal 1 or Goal 3, rather than reproducing the substantial Goal 2 share and choice dispersion observed in the data.

5.6 Gender Differences in Structural Parameters

Women are substantially more likely than men to choose conservatively under the benchmark model. As a descriptive accounting exercise, we estimate each model allowing only its focal parameter to vary by gender while holding the remaining parameters common. The EU-based models generate little gender variation in their fitted parameters and reproduce 28 percent of the observed choice gap. Loss aversion assigns women a larger fitted coefficient than men (2.50 versus 1.76) and reproduces 55 percent of the gap, while PCN assigns women greater fitted contingency neglect (1.00 versus 0.68) and reproduces 60 percent. Gender-specific PCN can therefore organize much of the field difference in choice, though loss aversion provides a similar accounting. While the exercise does not establish a gender difference in the underlying behavioral primitive, Section 6 presents direct experimental evidence on this question.

6 EXPERIMENTAL VALIDATION OF MECHANISMS

The field evidence shows that pairwise contingency neglect (PCN) fits observed choices better than the standard model and other prominent alternatives, with more mixed evidence for gain-loss utility. Model fit alone, however, cannot establish that the proposed mechanism is operative—a model may predict choices well without describing the process that generates them. This section therefore draws on three experiments with complementary roles. Experiment B (Section 6.1) uses repeated, incentivized goal choices across six menus to test whether stable person-level primitives such as risk or loss aversion organize behavior, or whether PCN better predicts how choices shift as the menu changes. Experiments C

and A (Section 6.2) target the mechanism directly: they measure the proposed inferential distortion at the level of individual beliefs and test whether choice responds to alternative presentations of likelihood information. Within Experiment A, the fourth condition (RQ-IP), which displayed accurate pairwise conditional probabilities, provides a preregistered, incentivized replication of the central presentation test in a larger independent sample. Because Experiment A was conducted on Prolific in 2026, it also establishes that the central result from the earlier MTurk experiments survives on a different platform under real financial incentives.

Experiment A's first three conditions, reported in Section 3, establish that conservative, heterogeneous, and gendered choice persists when the GQ menu is translated into an explicit lottery and paired with real financial consequences. This validation ladder supports using Experiment C's hypothetical GQ menu as a diagnostic environment in which beliefs and decision processes can be measured directly. Complete instructions, selected screenshots, incentive details, comprehension checks, exclusion rules, and timelines appear in the Experimental Appendix; Appendix Table A13 summarizes each experiment's platform, date, sample size, arms, and incentives.

6.1 Predicting Repeated Goal Choices Across Menus (Experiment B)

Design. We administered Experiment B in May 2019 to employed U.S. adults aged 25–55 recruited through Amazon Mechanical Turk. Participants completed a short, incentive-compatible puzzle task framed as a GoalQuest-style goal-reward problem. After practice and a comprehension test, they learned that they would have four minutes to solve number grids and selected goals from six menus presented in succession, with one menu randomly selected for payment. The baseline menu resembled the field setting in attainment difficulty and offered nonlinearly increasing rewards; the remaining menus varied overall difficulty, the relative generosity of the high reward, or the number of options. We also elicited beliefs about goal attainment, an independent measure of loss aversion from hypothetical gambles, and the measures of relative ability and taste for competition associated with contextual sorting accounts. The final sample contains 277 participants making 1,662 incentivized choices.

Held-out prediction. The baseline menu reproduces the central field patterns in goal choice, attainment beliefs, overconfidence, and choice relative to subjective expected value. More important, the six-menu design permits a true predictive test: whether a model fitted to five of a participant's choices predicts the omitted choice. Table 6 reports this test for PCN, fitted loss aversion, and subjective expected utility, giving each model one fitted participant-specific parameter.

Table 6.

Experiment B - Held-Out Predictions Across Models

Held-Out Metric	Subjective EU	Pairwise CN	Loss Aversion	Difference (PCN minus LA)	p-value
Observed Choice Predicted to be Optimal	0.512	0.670	0.610	0.060 (0.021)	0.0043
Support Across Equally Best-Fitting Parameter Values	0.512	0.661	0.611	0.051 (0.020)	0.0129

Notes: This table reports leave-one-menu-out predictions in Experiment B. The sample contains 277 participants making six choices each, yielding 1,662 held-out participant-menu folds. For each participant and omitted menu, each model is fitted on the other five menus and used to predict the omitted choice. Subjective EU selects CARA curvature rho from [0, 0.001] in increments of 0.0001. Pairwise CN selects theta from [0, 1] in increments of 0.05, with rho fixed at 0.0003. Loss Aversion selects lambda from [1, 2.5] in increments of 0.05, with alpha fixed at 0.88. All three models use subjective attainment beliefs and fit one participant-specific parameter. The training-optimal set contains all grid values maximizing the number of training choices rationalized. The center-parameter hit uses the arithmetic mean of this set, and the stability score is the share of equally best-fitting values under which the held-out choice is optimal. Difference denotes Pairwise CN minus Loss Aversion. Standard errors and two-sided p-values are based on paired fold-level differences, with clustering by participant. Pairwise CN's results are unchanged when rho is instead fixed at 0.001.

PCN predicts 67.0 percent of omitted-menu choices, compared with 61.0 percent for fitted loss aversion and 51.2 percent for subjective expected utility. PCN's advantages are 6.0 percentage points over loss aversion ($p = 0.004$) and 15.8 points over subjective expected utility ($p < 0.001$); loss aversion also outperforms subjective expected utility by 9.7 points ($p < 0.001$). The ranking is unchanged when equally best-fitting parameter values are considered. The comparison is generous to the alternatives: loss aversion uses the participant-specific value that best fits the five training menus rather than the independently elicited measure, while subjective expected utility similarly selects its best-fitting curvature. All models are evaluated only on choices not used to select their parameters.

Broader descriptive comparisons. Appendix Table A14 reports the in-sample comparison across subjective expected utility, nonlinear probability weighting, fitted loss aversion, independently elicited loss aversion, PCN, relative ability, and taste for competition. PCN rationalizes the largest share of participants' repeated choices at every reported threshold. A single PCN parameter rationalizes at least five of a participant's six choices for 56.7 percent of participants, compared with 43.0 percent for fitted loss aversion and 25.3 percent when the loss-aversion parameter is instead fixed at the independently elicited value. Subjective expected utility and nonlinear probability weighting rationalize 28.9 and 34.7 percent, respectively, and the relative-ability and competition heuristics each rationalize fewer than 9 percent. Because Appendix Table A14 fits and evaluates each model on the same six choices and allows the models different degrees of flexibility, we treat it as descriptive evidence. Its most informative result is how much ground loss aversion loses when the freely fitted parameter is replaced by the participant's independently elicited measure. Together, Table 6 and Appendix Table A14 show that PCN predicts behavior across menus better than fitted loss aversion and that independently measured loss attitudes and contextual sorting measures provide comparatively little explanatory power.

6.2 Direct Evidence on Pairwise Contingency Neglect (Experiments C and A)

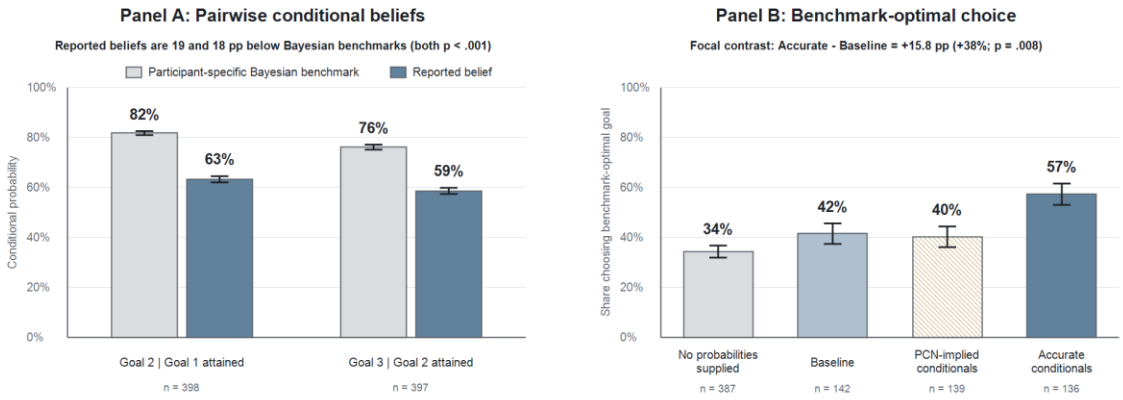
The field estimates and Experiment B establish PCN's predictive advantage; Experiments C and A test its mechanism directly. Both elicit conditional beliefs, randomize accurate conditional information while holding the underlying lottery fixed, and measure participants' comparison strategies. Experiment C additionally compares choices made with no supplied probabilities to choices under the biased conditionals implied by PCN, and Experiment A provides a preregistered, incentivized replication of the informationally equivalent presentation test.

Experiment C design. We administered Experiment C in July 2022 to 927 employed U.S. adults recruited through Amazon Mechanical Turk. The analytic sample comprises the 815 participants who passed all study screens, including a comprehension check; because screening preceded random assignment, exclusions were determined before treatment. Participants were randomly assigned to a belief-and-process module (398 participants) or a presentation module (417 participants). Both modules used a hypothetical GoalQuest menu with goals of 105, 110, and 115 units and rewards of $150, $450, and $900 (identical to Experiment A).

The belief-and-process module elicited each participant's marginal probabilities of attaining the three goals and, in randomized order, the conditional probabilities of attaining a higher goal given attainment of the lower goal. This design allows us to compare each participant's reported conditional belief with the Bayesian conditional implied by that participant's own marginal beliefs. Participants also reported which options they compared while deciding, and an auxiliary weather-judgment task examines whether the same conditional-inference error appears outside goal choice.

The presentation module presented the same menu under one of three randomly assigned probability displays. The baseline displayed the three attainment probabilities in marginal form. The accurate-conditional presentation—informationally equivalent to the baseline—instead displayed the probability of attaining Goal 1 together with the accurate probabilities of attaining Goal 2 conditional on Goal 1 and Goal 3 conditional on Goal 2. A third display paired Goal 1's attainment probability with the lower conditional probabilities implied by PCN; this biased-conditional presentation is deliberately not informationally equivalent to the first two and serves as a diagnostic comparison. Finally, because assignment across modules was itself random, choices in the belief-and-process module—where no probabilities were supplied—provide a no-information comparison in which participants had to construct the relevant probabilities themselves from a list of fictional figures estimated from empirical averages.

Figure 3. Conditional Beliefs and Choice in Experiment C



Notes: Panel A compares reported pairwise conditionals with participant-specific Bayesian benchmarks implied by marginal beliefs. Panel B plots benchmark-optimal choice; Baseline and Accurate are informationally equivalent, PCN-implied is diagnostic, and No probabilities supplied comes from the separately randomized belief-and-process module. Error bars show ±1 SE.

Conditional-belief evidence from Experiment C. Reported conditional judgments are anchored in participants' marginal beliefs (Appendix Table A15). As shown in Figure 3, Panel A, the mean reported probability of attaining Goal 2 conditional on attaining Goal 1 is 63 percent, against a mean participant-specific Bayesian benchmark of 82 percent; the corresponding means for attaining Goal 3 conditional on attaining Goal 2 are 59 and 76 percent. Reported conditionals therefore fall below their Bayesian benchmarks by 19 and 18 percentage points, respectively ($p < 0.001$ for both). To our knowledge, this is the first systematic evidence that people understate pairwise conditional probabilities relative to the Bayesian conditionals implied by their own marginal beliefs.[14]

The errors are systematically related to participants' own marginal beliefs. When we regress each reported conditional on the respondent-specific Bayesian conditional and the marginal probability of attaining the higher goal, the marginal probability receives a coefficient of 0.526 (SE 0.053; Appendix Table A15). Thus, even after accounting for the Bayesian benchmark implied by a participant's beliefs, the reported conditional is pulled toward the lower unconditional probability of attaining the higher goal—the pattern predicted by PCN. This participant-specific marginal anchor also outperforms generic compression toward 0.50 in held-out prediction. Across respondent-grouped cross-validation splits, the one-parameter PCN model yields an RMSE of 0.198, compared with 0.211 for an equally parsimonious 0.50-compression model, and performs better in all 20 splits.

Fitting the one-parameter PCN restriction to the Experiment C belief reports yields $\theta = 0.816$, nearly identical to the field estimate of $\theta = 0.812$ under the common low-curvature calibration. Because

[14] Prior research documents partition dependence and other failures of contingent reasoning, but does not, to our knowledge, establish this directional underestimation against respondent-specific Bayesian benchmarks (Fox and Clemen, 2005; Martínez-Marquina, Niederle, and Vespa, 2019; Rees-Jones, Shorrer, and Tergiman, 2024).

the experimental estimate is obtained from reported beliefs whereas the field estimate is obtained from choices, we view this correspondence as suggestive rather than as evidence of exact parameter stability.

Conditional-belief errors are also strongly related to choice. A 10-percentage-point larger error in the reported Goal 3 conditional is associated with a 7.7-percentage-point reduction in benchmark-optimal choice after controlling for participants' marginal beliefs ($p < 0.001$). The auxiliary weather task produces the same qualitative pattern of systematic conditional-probability understatement, showing that the phenomenon extends beyond goal-choice menus.

Conditional-belief evidence from Experiment A. Experiment A reproduces the belief distortion in an independent sample in which probabilities were externally supplied, so the correct conditional values—89.2 percent for attaining Option 2 conditional on attaining Option 1 and 87.8 percent for attaining Option 3 conditional on attaining Option 2—are objective benchmarks rather than constructions from participants' own beliefs. In each of the three conditions displaying probabilities in marginal form, reported conditional beliefs fell substantially below these values, and the distortion appears alike in the GoalQuest frame, the hypothetical lottery frame, and the incentivized lottery frame. In the incentivized marginal-probability condition (RQ-I), mean conditional reports were 57.8 and 46.1 percent. Providing the two accurate conditionals in the informationally equivalent pairwise condition (RQ-IP) raised the corresponding reports to 76.1 and 70.6 percent and reduced average absolute error from 36.7 to 15.5 percentage points—a substantial, though incomplete, correction.

Belief errors again predict choice. Averaging the condition-specific estimates across Experiment A's three marginal-probability conditions, a 10-percentage-point increase in absolute conditional-belief error predicts a 7.0-percentage-point increase in conservative choice, and the relationship is statistically significant in each condition ($p < 0.001$). Together with Experiment C, these findings connect conditional-probability understatement to conservative choice.

Causal evidence from Experiment C. Figure 3, Panel B summarizes the randomized presentation results. Under the informationally equivalent baseline, 41.5 percent of participants selected the expected-value-maximizing goal; under the accurate-conditional presentation, 57.4 percent did. Supplying accurate conditionals therefore increased benchmark-optimal choice by 15.8 percentage points, or 38 percent relative to baseline (95 percent CI: 4.1–27.5; $p = 0.008$).

The remaining comparisons form a diagnostic pattern. When shown the biased conditionals implied by PCN, 40.3 percent of participants selected the benchmark-optimal goal, statistically indistinguishable from the 41.5 percent under the baseline presentation ($p = 0.83$). When no probabilities were supplied, 34.4 percent did so, also not statistically distinguishable from choice under the PCN-implied conditionals ($p = 0.22$). These comparisons are not precise enough to establish exact equivalence, but their configuration is informative. When participants receive no probability information and must

construct the relevant probabilities themselves, their choices resemble those made when the experiment explicitly supplies the biased conditionals implied by PCN; making those biased conditionals explicit changes little; and replacing them with accurate conditionals shifts choice substantially toward the benchmark-optimal goal. The no-information comparison therefore supports the interpretation that the PCN-implied presentation approximates the conditional probabilities participants naturally construct.

Causal evidence from Experiment A. Experiment A replicates the informationally equivalent comparison with preregistration and real incentives. Among participants facing the incentivized lottery with probabilities displayed in marginal form, 32.2 percent selected the expected-value-maximizing option. Displaying the same objective probability distribution in accurate pairwise conditional form raised this share to 47.4 percent (Figure 1, Panel B)—an increase of 15.3 percentage points (95 percent CI: 7.3–23.3; $p < 0.001$), or 47 percent relative to baseline. The accompanying movement in reported conditional beliefs is consistent with the proposed channel.[15]

Relative to their respective baselines, accurate conditionals increased benchmark-optimal choice by 38 percent in Experiment C and 47 percent in Experiment A. The replication recovers both the direction and the economic importance of the effect. Because the underlying lottery is unchanged within each comparison, stable preferences, loss aversion, utility curvature, and baseline beliefs cannot by themselves explain the response to the alternative presentation. Making the conditional probabilities explicit may also focus attention or simplify computation. The belief results show, however, that correcting marginally anchored conditional judgments is an important part of the response.

Self-reported comparison processes. Participants' descriptions of how they chose provide supporting evidence for pairwise evaluation. In Experiment C, 86.2 percent reported making at least one pairwise comparison, with adjacent comparisons especially common. Experiment A yields the same pattern: across the three marginal-probability conditions, 85.5 percent reported comparing Options 1 and 2, and 87.2 percent reported comparing Options 2 and 3. Providing accurate conditionals in Experiment A did not measurably change reported comparison strategies, suggesting that the intervention changed the information entering existing pairwise comparisons rather than the comparison rule itself. These reports support pairwise evaluation but do not identify its direction or stopping rule.

Gender. The experimental evidence does not yield a uniform gender pattern. Experiment B's repeated-choice comparisons and the direction of Experiment C's choice response both point to a greater

[15]An exploratory product-of-coefficients calculation combines the randomized effect of displaying accurate conditionals on belief error with the belief-error–choice relationship estimated only in the three marginal-probability conditions, where participants had to construct the conditionals themselves. The implied indirect reduction in conservative choice is 10.4 percentage points (95% bootstrap CI: 7.6–13.4), approximately 83% of the adjusted 12.5-percentage-point difference between RQ-IP and the pooled marginal-probability conditions. Because beliefs were elicited after choice, we interpret this estimate as suggestive evidence.

role for PCN among women, consistent with the gender difference observed in the field. In contrast, Experiment C's belief-based estimates of theta are nearly identical for women and men, and Experiment A's treatment effect is directionally larger among men. The treatment-by-gender interactions are imprecisely estimated, so the experiments neither establish nor rule out meaningful gender heterogeneity in PCN. Taken together, the evidence is supportive in some designs but does not provide a consistent experimental explanation for the field gender gap.

Collectively, Experiments C and A provide direct and complementary evidence for pairwise contingency neglect. Participants systematically underestimate pairwise conditional probabilities in both goal-choice and weather judgments, and anchoring on their own marginal beliefs accounts for a substantial component of the error. When participants receive no explicit probability information, their choices resemble those made when the experiment supplies the biased conditionals implied by PCN, and in two independent informationally equivalent comparisons, supplying accurate conditional probabilities substantially increases benchmark-optimal choice. Self-reports further show that most participants compare adjacent options. The convergence of belief, choice, and process evidence supports PCN as a causally consequential account of choice in nested menus.

7 PAIRWISE CONTINGENCY NEGLECT IN INSURANCE CHOICE

The mechanism we document in GoalQuest is not specific to workplace goals. It should operate in any menu of nested lotteries—that is, in any setting where comparing adjacent options requires judging the probability of a more extreme outcome conditional on a less extreme one. Portfolio choices with layered risk exposure, contingent contracts, and tiered bonus schemes share this structure. Insurance menus ordered by coverage are perhaps the most direct case. Upgrading to more generous coverage costs an additional premium, while the benefit—reduced out-of-pocket spending—arrives only if a loss occurs and generally grows with its severity. Just as the value of a higher goal depends on the probability of attaining it conditional on attaining the lower one, the value of more generous coverage depends on the probability of a severe loss conditional on experiencing any loss at all.

Insurance differs from goal choice in one important respect: the premium is paid whether or not a loss occurs. In GoalQuest, an employee who misses the lower goal receives nothing under either option, so the probability of reaching the lower goal scales both options equally and cannot change which is preferred. In insurance, by contrast, a premium increase incurred in every state must be weighed against benefits received only in loss states. Two probabilities therefore shape the comparison between adjacent plans: the probability of experiencing any loss, and the probability of a severe loss conditional on experiencing one. Consumers arrive at these two probabilities in different ways. The conditional probability is defined by the cost-sharing terms of the specific plans on offer, so ordinary experience cannot supply it and consumers

must construct it. The probability of any loss depends on the risk: consumers can recall it from repeated experience when losses are frequent but must construct it when losses are rare.

Which probabilities must be constructed determines the direction of the bias. For frequent risks, such as prescription-drug spending, consumers can recall how often they incur expenses but must still construct the conditional probability of reaching the severe tail. Contingency neglect pulls that constructed probability downward, so consumers understate the value of additional coverage, and the model predicts underinsurance. For rare risks, such as major home damage, both probabilities must be constructed. Consumers who evaluate coverage from within the loss scenario fail to discount its benefits fully by the small chance that the scenario occurs, and the model predicts overinsurance. The frequency of the underlying risk thus determines whether PCN shifts demand toward less or more insurance.

The framework describes how consumers choose among ordered coverage levels once they engage with the insurance decision (this does not preclude the menu including no insurance as its lowest-coverage option). It does not explain whether consumers notice, consider, or engage with the insurance decision in the first place, distinguishing it from evidence on the neglect of rare-event coverage, annuitization, and long-term-care insurance.[16] Within this scope, the framework helps organize two prominent patterns: low insurance demand in prescription-drug plan choice (Abaluck and Gruber, 2011; Heiss et al., 2013) and the unusually strong demand for low deductibles against modest, low-probability losses (Sydnor, 2010; Barseghyan et al., 2013). It also yields a prediction that stable preferences cannot— holding the objective lottery fixed, restating risk in terms of accurate conditional probabilities should change coverage choice.

7.1 A Framework for Pairwise Insurance Choice

Setup. Consider a consumer choosing between two plans, Plans 1 and 2. Plan 2 charges a higher premium in every state but pays an incremental indemnity that rises with the realized loss. The plans' cost-sharing thresholds divide possible losses into three regions: a low-loss region in which the upgrade provides little or no additional indemnity, a moderate-loss region in which it provides a partial benefit, and a high-loss region in which its benefit is largest. Deductible and coinsurance contracts both share this structure; their cost-sharing terms determine the size of the resulting bias.[17]

[16] Research on rare-event insurance emphasizes non-consideration, episodic attention, and learning after disasters (Kunreuther et al., 1978; Camerer and Kunreuther, 1989; Gallagher, 2014). Annuity and long-term-care puzzles likewise concern entry into a market rather than an adjacent comparison among coverage levels (Yaari, 1965; Benartzi, Previtero, and Thaler, 2011; Brown and Finkelstein, 2009). Luttmer, de Oliveira, and Taubinsky (2026) identify a different failure of contingent reasoning in annuitization: under-appreciation of how annuitization interacts with saving decisions.

[17] Application requires that relevant nonfinancial differences across plans be negligible or incorporated into the state-specific utilities. Multiple cost-sharing features can be accommodated as long as the incremental utilities remain ordered. Under coinsurance, the incremental benefit of coverage continues to grow in the severe tail, making H larger and strengthening the underinsurance effect. A deductible caps the incremental benefit at the deductible difference, making H smaller and the underinsurance effect weaker. Thus, contract type affects the magnitude rather than the sign of the bias.

Let U_L, U_M, and U_H denote the incremental utility of Plan 2 over Plan 1 in each region, incorporating both the premium difference and the incremental indemnity, with $U_H > U_M > U_L$. Define $A = U_M - U_L$ and $H = U_H - U_M$. Let a be the probability that a loss occurs for which the upgrade provides some benefit, and let q be the probability of the high-loss region conditional on such a loss, so that the unconditional probability of a high loss is $s = aq$.

A consumer who evaluates the plans pairwise values the upgrade according to:

$$V = (1-a)U_L + a(1-q)U_M + aqU_H = U_L + a(A + qH) \tag{7.1}$$

The contract makes the two probabilities enter equation (7.1) asymmetrically: a multiplies both incremental benefits, while q multiplies only the additional benefit received in the high-loss region.

Recalled and constructed probabilities. The two probabilities differ in how consumers naturally obtain them. The conditional probability q is defined by the plans being compared, so it must ordinarily be constructed. The probability of any loss a can instead be recalled from repeated experience when losses are frequent. When losses are rare, personal experience provides few relevant observations, so *a* must ordinarily be constructed as well. When a consumer must construct a probability rather than recall it, we assume she anchors on a readily available value and adjusts incompletely. A single parameter θ between zero and one measures the extent of this failure to adjust. When the probability of any loss must be constructed, $\tilde{a} = (1-\theta)a + \theta$; when it is recalled and incorporated directly, $\tilde{a} = a$. When the conditional probability of a severe loss must be constructed, $\tilde{q} = (1-\theta)q + \theta s$; when it is supplied directly and incorporated, $\tilde{q} = q$. We refer to these relationships collectively as equation (7.2).

The two errors work in opposite directions. Because s < q, a constructed conditional probability is anchored toward the lower unconditional probability of a high loss and is therefore too low—the same error that generates conservative goal choice. Because $a < 1$, a constructed probability of any loss is biased upward toward the focal loss scenario. This dependence on which hypothesis the comparison makes focal is consistent with support theory and evidence on partition priming (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003). The same θ governs both adjustments; experience determines whether the outer probability is recalled or constructed and hence whether only the inner distortion or both distortions operate.

Implications for insurance demand. Let $\tilde{V}$ denote the perceived value of the upgrade—equation (7.1) evaluated at the perceived probabilities—and let $B = \tilde{V} - V$ denote the resulting bias. Because the conditional probability must ordinarily be constructed, the relevant distinction is whether the probability of any loss can be recalled from experience or must be constructed as well. Substituting equations yields:

$$\begin{aligned} B_{\text{any-loss probability recalled}} &= -\theta(1-a)aqH < 0 \\ B_{\text{any-loss probability constructed}} &= \theta(1-a)A + \theta(1-\theta)(1\text{-}a)^2 qH \geq 0 \end{aligned} \tag{7.3}$$

When the probability of any loss is recalled from experience, the only error is an understated conditional probability of a severe loss. Additional coverage then appears less valuable than it is, and PCN produces underinsurance. When the probability of any loss must also be constructed, both terms in the second line of equation (7.3) are nonnegative, and PCN produces overinsurance.[18] Underinsurance therefore requires a particular configuration: the probability of any loss is recalled, while the conditional probability of a severe loss must still be constructed.

Experience and the probability of any loss. Whether a consumer recalls or constructs the probability of any loss depends on her experience with the risk. Let $n = aT$ denote the expected number of exposure periods with at least one relevant loss event over horizon T. Among retirement-age adults, the annual probability of any prescription use is close to 90 percent (NCHS 2023), so twenty years of exposure imply roughly eighteen years with at least one prescription-drug expenditure—and typically far more individual spending events. By contrast, the 4.2 percent annual homeowners-claim rate in Sydnor (2010) implies fewer than one expected claim over the same horizon. Repeated experience can therefore supply the probability of any prescription spending, whereas personal claims history ordinarily cannot supply the probability of home damage. The empirical rule is asymmetric. Because insurance commonly covers rare losses whose probabilities cannot be learned reliably from experience, PCN predicts overinsurance as the default. Underinsurance is the narrower case, arising for frequent risks when the value of additional coverage is concentrated in the severe tail. To give a sense of magnitudes, the model reverses the benchmark ranking of Sydnor's (2010) lower deductible at levels of θ below those estimated in the field.[19]

7.2 Experimental Evidence on Insurance Choice

We test the framework in two scenario-based experiments that use different forms of equivalence to make departures from benchmark-optimal choice interpretable. Experiment D studies frequent prescription-drug spending using informationally equivalent presentations, each encoding the same expenditure distribution; PCN predicts too little demand for the benchmark-optimal generous plan. Experiment E studies rare home damage using decision-equivalent presentations that disclose different risk information but leave the same plan benchmark-optimal over broad ranges of loss beliefs and risk aversion. PCN predicts excess coverage and a directional response to whether the presentation emphasizes the loss or no-loss state. The Experimental Appendix summarizes designs and provides screenshots.

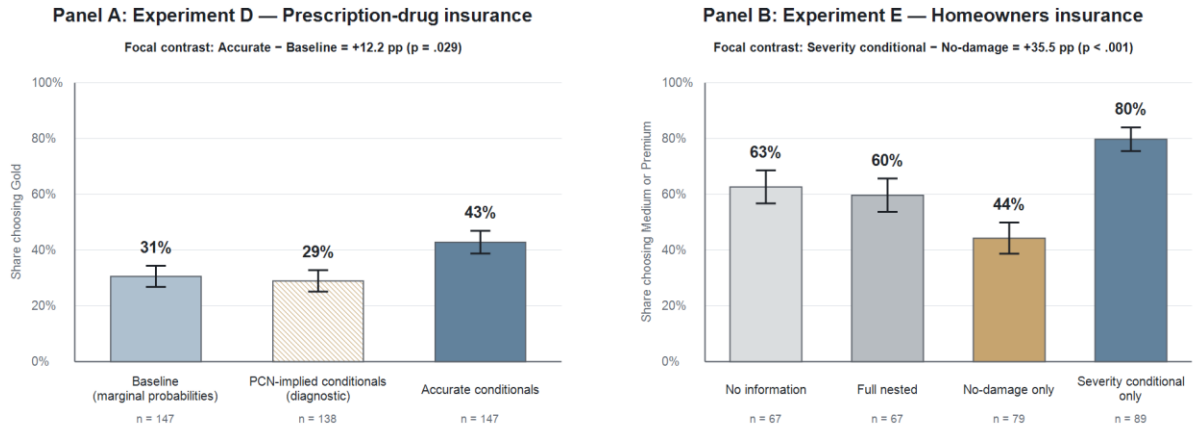
---

[18] With a common adjustment parameter, $\tilde{a} = a + \theta(1-a)$ and $\tilde{q} = q[1-\theta(1-a)]$, so $\tilde{a}\tilde{q} - aq = \theta(1-\theta)(1\text{-}a)^2 q \geq 0$ and $\tilde{a} - a = \theta(1-a) > 0$. Both components are therefore nonnegative. At complete anchoring, the severity component returns to zero while the moderate-region component remains positive.

[19] In Sydnor (2010), customers choosing the $500 deductible paid an average incremental premium of $99.85 for an average expected benefit of $19.93. Under equation (7.2) the upgrade becomes attractive once â/a exceeds about 5, which at a 4.2 percent claim rate requires θ of roughly 0.18 — well below our field estimate. The severe-tail correction is small for a deductible contract, where the incremental benefit is capped at the difference in deductibles.

Prescription-drug insurance: (Experiment D). Experiment D analyzes 432 U.S. adults recruited through Amazon Mechanical Turk in 2023. After instructions and comprehension questions, participants were presented a hypothetical insurance scenario and chose among no insurance, a Silver Plan, and a more generous Gold Plan. Silver and Gold differed in premiums and coinsurance, so the incremental value of Gold rose with prescription-drug spending. Under both expenditure distributions used in the experiment, Gold minimized expected cost and remained optimal under CARA expected utility for any $0 \leq \rho$.

Figure 4. Insurance Choice under Informationally Equivalent and Decision-Equivalent Frames



Notes: Bars show sample shares on a common 0–100% scale; error bars show ±1 SE. In Experiment D, Baseline and Accurate encode the same objective expenditure distribution, while PCN-implied is diagnostic. In Experiment E, disclosed probability components differ across frames while Basic remains benchmark-optimal over the maintained domain.

In the Baseline arm, participants learned that annual expenditures would exceed $1,280 with 60 percent probability and $1,657 with 40 percent probability. Comparing Gold with Silver therefore required a conditional inference: among the cases exceeding $1,280, two-thirds also exceed $1,657. The Accurate Conditionals arm stated this 67 percent conditional probability directly while preserving the same objective expenditure distribution, so Baseline versus Accurate Conditionals is an informationally equivalent comparison. The PCN-Implied Conditionals arm used the same conditional format but displayed 40 percent, the value implied by complete contingency neglect; Accurate versus PCN-Implied thus holds the format fixed and varies only the displayed probability, while Baseline versus PCN-Implied compares the probability participants construct on their own with the value complete neglect predicts. All three arms also stated that Gold would be the least expensive plan if spending crossed the upper threshold.

Figure 4, Panel A, reports all three comparisons. Gold choice rose from 30.6 percent in Baseline to 42.9 percent under Accurate Conditionals—a 12.2-percentage-point increase from re-expressing the same objective information in conditional form ($p = 0.029$). Gold choice was 29.0 percent under PCN-Implied Conditionals, 13.9 points below Accurate Conditionals ($p = 0.015$) and 1.6 points below Baseline ($p = 0.82$). The similarity between the Baseline and PCN-Implied arms is consistent with participants constructing a

conditional probability close to the fully neglected value although a null cannot establish equivalence. The Baseline–Accurate comparison shows that expressing the same risk information in the form the pairwise comparison requires changes plan choice; the design does not distinguish whether the response operates through beliefs, attention, or ease of calculation. The PCN-Implied arm provides a complementary diagnostic, showing how participants choose when the displayed conditional probability equals the value predicted by complete contingency neglect.

Homeowners insurance: (Experiment E). Experiment E places participants in a hypothetical home-insurance scenario with a low baseline probability of loss. We analyze 302 responses from U.S. adults recruited through Amazon Mechanical Turk on October 22, 2022. Participants were asked to choose one of three homeowners-insurance plans adapted from Sydnor (2010). The plans were described as lender-mandated, so that participants chose how much coverage to buy, not whether to insure. The Basic Plan charged a $616 premium with a $1,000 deductible, the Medium Plan $716 with a $500 deductible, and the Premium Plan $803 with a $250 deductible. The outcome of interest is the choice of more generous coverage than Basic—that is, Medium or Premium.

Participants were randomly assigned to one of four presentations of the damage risk. The presentations are decision-equivalent in that each discloses different components of the risk, but the same plan remains benchmark-optimal under all of them for broad ranges of preferences and beliefs. Baseline (labeled No information in Figure 4) displayed only the plan menu. Full Nested reported a 4 percent probability of any damage and, conditional on damage, a 75 percent probability of severe damage (exceeding $2,500). No-damage only reported the complementary 96 percent probability of no damage. The Severity conditional only reported just the 75 percent probability of severe damage conditional on damage. Under every presentation, Basic minimized expected cost whenever the probability of any damage was at most 20 percent and remained optimal under CARA expected utility with $0 \leq \rho \leq 0.001$ up to 10.5 percent (more than twice the actuarial rate in the data).

The presentations were designed to vary whether participants adequately incorporate the low probability of damage. We treated Baseline as representative of how consumers ordinarily approach a low-probability insurance decision and, following the framework, expected participants to give too little weight to how rarely damage occurs. Stating the 96 percent probability of no damage makes that rarity explicit and because the higher premium is paid in cases without damage, we presumed this presentation should reduce the perceived value of additional coverage and increase Basic Plan choice. Presenting only severity conditional on damage should instead focus the comparison inside the damage scenario, further crowding out the low probability of reaching it, and reduce Basic Plan choice. We therefore predicted that Basic Plan choice would be greatest under No-damage only, intermediate under Baseline, and lowest under Severity

conditional only. We were agnostic as to the position of Full Nested, which supplies both probabilities, in this ordering as it served as a complete-information reference.

Figure 4, Panel B, reports the results, which match the predicted ordering: Basic Plan choice was greatest under No-damage only, intermediate under Baseline, and lowest under Severity conditional only. The share choosing more generous coverage differed by 35.5 percentage points between the Severity conditional only and No-damage only presentations ($p < 0.001$), while choices under Full Nested were close to Baseline. Because Basic remained benchmark-optimal throughout the reasonable range of damage risk and risk aversion, these presentations shifted choices without changing the benchmark recommendation.

Taken together, Experiments D and E test the two distortions the framework identifies. In Experiment D, where the probability of any spending can be recalled and the conditional error operates alone, most participants chose less coverage than the benchmark, and supplying the conditional tail probability in informationally equivalent form moved choice toward it. In Experiment E, where both probabilities must be constructed, most participants chose more coverage than the benchmark, and decision-equivalent presentations that shifted attention between the loss scenario and its complement moved choice in the directions the framework implies.

8 CONCLUSION

We study financial risk taking using decisions, productivity, and contemporaneous beliefs from more than 20,000 employees across 34 iterations of a large all-or-nothing goal-reward program. The setting offers a rare combination: a simple lottery structure, meaningful and varied stakes, broad worker diversity, near-universal participation, and direct measures of perceived risk at the moment of choice. We document substantial conservative choice and excess heterogeneity, with large implied reward losses. Conservative choice is particularly pronounced among women, and both conservatism and heterogeneity persist in complementary experiments that reduce field-specific confounds. In a structural horse race, standard utility-based risk preferences, biased beliefs, and probability weighting do not account well for the data. Loss aversion performs better, but its explanatory power weakens when disciplined by independent measurement.

The strongest overall support is for pairwise contingency neglect, under which employees evaluate nested options through adjacent comparisons and insufficiently distinguish conditional from marginal likelihoods. This mechanism accounts for conservative choice and excess heterogeneity better than competing benchmarks in both field and laboratory data, predicts behavior in held-out programs, and is supported by process, correlational, and causal evidence. In field estimates, greater fitted contingency neglect among women descriptively accounts for a substantial share of the gender gap in conservative choice, though the experiments provide only mixed support for this channel.

The insurance experiments demonstrate the broader relevance of the mechanism. Equivalent representations of the same risks, preserving either the underlying information or the benchmark-optimal choice, produce systematically different plan choices, violating descriptive invariance. The direction of these effects is consistent with field evidence of underinsurance in prescription-drug coverage and of excessive coverage against modest homeowners risks. Pairwise contingency neglect may therefore help explain insurance anomalies that otherwise appear unrelated or require distinct behavioral accounts.

More broadly, the results suggest that understanding the process generating risky choice is essential for interpreting observed behavior, drawing welfare conclusions, and designing contracts. If departures from benchmark choice partly reflect a correctable inferential error rather than stable risk preferences, standard welfare analysis may be misleading. Effective contract design should therefore attend not only to prices and incentives, but also to representations that make the relevant conditional tradeoffs transparent.

9 REFERENCES

Abaluck, J., & J. Gruber. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program. *American Economic Review*, 101(4): 1180–1210.

Ahn, D., & H. Ergin. (2010). Framing contingencies. *Econometrica*, 78(2): 655–695.

Alaoui, L., & A. Penta. (2025). What's in a u? Economics Working Paper Series No. 1909, Universitat Pompeu Fabra.

Barseghyan, L., F. Molinari, T. O'Donoghue, & J. C. Teitelbaum. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, 103(6): 2499–2529.

Barseghyan, L., F. Molinari, T. O'Donoghue, & J. C. Teitelbaum. (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, 56(2): 501–564.

Benartzi, S., & R. H. Thaler. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives*, 21(3): 81–104.

Benartzi, S., A. Previtero, & R. H. Thaler. (2011). Annuitization puzzles. *Journal of Economic Perspectives*, 25(4): 143–164.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 2 (pp. 69–186). Amsterdam: North-Holland.

Bhargava, S., G. Loewenstein, & J. Sydnor. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3): 1319–1372.

Bordalo, P., N. Gennaioli, & A. Shleifer. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3): 1243–1285.

Bordalo, P., N. Gennaioli, & A. Shleifer. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5): 803–843.

Brown, J. R., & A. Finkelstein. (2009). The private market for long-term care insurance in the United States: A review of the evidence. *Journal of Risk and Insurance*, 76(1): 5–29.

Camerer, C. F., & H. Kunreuther. (1989). Decision processes for low probability events: Policy implications. *Journal of Policy Analysis and Management*, 8(4): 565–592.

Christenfeld, N. (1995). Choices from identical options. *Psychological Science*, 6(1): 50–55.

Enke, B., & T. Graeber. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4): 2021–2067.

Ericson, K., & A. Starc. (2012). Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts Connector. *American Economic Review*, 102(3): 493–497.

Esponda, I., & E. Vespa. (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4): 180–202.

Fox, C. R., & R. T. Clemen. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science*, 51(9): 1417–1432.
Fox, C. R., & Y. Rottenstreich. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, 14(3): 195–200.
Gallagher, J. (2014). Learning about an infrequent event: Evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics*, 6(3): 206–233.
Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3): 667–686.
Heath, C., R. P. Larrick, & G. Wu. (1999). Goals as reference points. *Cognitive Psychology*, 38(1): 79–109.
Heiss, F., A. Leive, D. McFadden, & J. Winter. (2013). Plan selection in Medicare Part D: Evidence from administrative data. *Journal of Health Economics*, 32(6): 1325–1344.
Holt, C. A., & S. K. Laury. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5): 1644–1655.
Jaspersen, J. G., M. A. Ragin, & J. R. Sydnor. (2022). Predicting insurance demand from risk attitudes. *Journal of Risk and Insurance*, 89: 63–96.
Kahneman, D., & A. Tversky. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–291.
Kunreuther, H., R. Ginsberg, L. Miller, P. Sagi, P. Slovic, B. Borkan, & N. Katz. (1978). *Disaster insurance protection: Public policy lessons*. New York: Wiley.
Kőszegi, B., & M. Rabin. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4): 1133–1165.
Kőszegi, B., & A. Szeidl. (2013). A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1): 53–104.
Kusev, P., H. Purser, R. Heilman, A. J. Cooke, P. Van Schaik, V. Baranova, R. Martin, & P. Ayton. (2017). Understanding risky behavior: The influence of cognitive, emotional, and hormonal factors on decision-making under risk. *Frontiers in Psychology*, 8: 102.
Loomes, G., & R. Sugden. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368): 805–824.
Luttmer, E. F. P., P. de Oliveira, & D. Taubinsky. (2026). Failures of contingent reasoning in annuitization decisions. *American Economic Journal: Microeconomics*, 18(3): 454–480.
Martínez-Marquina, A., M. Niederle, & E. Vespa. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109(10): 3437–3474.
National Center for Health Statistics. (2023). QuickStats: Percentage of adults aged ≥18 years who took prescription medication during the past 12 months, by sex and age group—National Health Interview Survey, United States, 2021. *Morbidity and Mortality Weekly Report*, 72(16): 450.
Niederle, M. (2017). Gender. In J. Kagel & A. Roth (Eds.), *The handbook of experimental economics*, Vol. 2 (pp. 481–562). Princeton, NJ: Princeton University Press.
Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3): 497–527.
Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5): 1281–1292.
Rees-Jones, A., R. Shorrer, & C. Tergiman. (2024). Correlation neglect in student-to-school matching. *American Economic Journal: Microeconomics*, 16(3): 1–42.
Simonson, I., & A. Tversky. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29(3): 281–295.
Sunstein, C. R., & R. J. Zeckhauser. (2010). Dreadful possibilities, neglected probabilities. In E. Michel-Kerjan & P. Slovic (Eds.), *The irrational economist: Making decisions in a dangerous world*. New York: PublicAffairs.
Sydnor, J. (2010). (Over)insuring modest risks. *American Economic Journal: Applied Economics*, 2(4): 177–199.
Tversky, A., & D. Kahneman. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4): 297–323.
Tversky, A., & D. J. Koehler. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101(4): 547–561.
Valenzuela, A. M., & P. Raghubir. (2009). Position-based beliefs: The center-stage effect. *Journal of Consumer Psychology*, 19(2): 185–196.
von Neumann, J., & O. Morgenstern. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.
Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, 32(2): 137–150.

Appendix A.1: Generalized Framework with Discretionary Effort

The theoretical framework in the main text abstracts away from effort motives, treating goal choice as a pure lottery selection. A natural concern is that conservative goal choice could reflect the avoidance of costly effort rather than attitudes toward risk. Here we generalize the baseline model to incorporate discretionary effort, derive the conditions for optimal goal choice in this richer setting, and show that the timing of our belief elicitation implies that the main-text estimates of conservative choice should be interpreted as a lower bound—that is, the simplified framework, if anything, understates the degree of conservatism.

Setup. Consider a risk-neutral employee who jointly selects a productivity goal and a level of costly, productivity-enhancing effort. As before, goal $G_n \in \{G_l, G_h\}$ yields reward x_n with probability $s_n(e)$ and zero otherwise, with $x_h > x_l$ and, for a given effort level, $s_h(e) < s_l(e)$. Goals are nested: attainment of G_h implies attainment of G_l. The employee chooses effort $e \geq 0$ at cost $c(e)$, where c is increasing and convex with $c(0) = 0$. Higher effort weakly increases the likelihood of attaining each goal: $s_n'(e) \geq 0$. Effort is committed at the time of goal selection and cannot be revised afterward. We continue to assume a discount rate of one.

Decision Rule. The employee jointly selects a goal and effort level to maximize:

$$\max_{n\in\{h,l\},\ e\geq 0} x_n\, \hat{s}_n(e) - c(e)$$

Let e_n^* denote optimal effort conditional on having chosen goal n, and let e^* denote optimal effort for the chosen goal. The employee selects the high goal if two conditions are jointly satisfied:

Condition 1 (Reward Favorability): Under optimal effort for the chosen goal, the reward advantage of the high goal must offset its lower likelihood:

$$\frac{x_h}{x_l} > \frac{\hat{s}_l(e^*)}{\hat{s}_h(e^*)}$$

Condition 2 (Effort-Inclusive Comparison): The expected value of the high goal under its own optimal effort must exceed the expected value of the low goal under the low goal's optimal effort, net of any difference in effort costs:

$$x_h\, \hat{s}_h(e_h^*) - c(e_h^*) > x_l\, \hat{s}_l(e_l^*) - c(e_l^*)$$

The first condition asks whether the high goal looks favorable holding effort fixed at the observed level. The second asks whether it remains favorable after accounting for the possibility that effort would be optimized differently under the alternative goal. Both must hold for the high goal to be truly optimal.

What We Observe. The critical institutional fact is that we elicit beliefs *after* goal selection. Employees report $\hat{s}_n(e^*)$—perceived likelihoods of attaining each goal given optimal effort under their *chosen* goal. This means we observe the inputs needed to evaluate Condition 1 but not Condition 2,

because we do not observe the counterfactual beliefs $\hat{s}_n(e^*_{n'})$ that would obtain if the employee had chosen the other goal and optimized effort accordingly.

Bounding Result. This asymmetry has directional implications for the characterization of choice. The following table considers the four possible scenarios defined by observed goal choice and whether Condition 1 is satisfied:

Observed Choice	Condition 1	Simplified Characterization	Possible True Characterization
G_h	Satisfied	Optimal	Could be aggressive (if Condition 2 fails)
G_h	Not satisfied	Aggressive	Aggressive (correctly identified)
G_l	Satisfied	Conservative	Conservative (correctly identified)
G_l	Not satisfied	Optimal	Could be conservative (if Condition 2 holds)

In scenarios (i) and (iv), the simplified framework may misclassify choices as optimal when they are in fact aggressive or conservative, respectively. Crucially, both types of error work in the same direction: they *overestimate* the share of optimal choice and *underestimate* departures from the benchmark. The main-text estimates of conservative choice should therefore be interpreted as a lower bound: if effort motives matter, the true share of conservative choice is at least as large as—and likely larger than—what we report.

Extension to Three Goals. The argument extends directly to the three-goal GoalQuest menu. If employees first compare Goals 1 and 2 and then compare the preferred option to Goal 3, the same logic applies at each stage: we observe beliefs conditioned on optimal effort for the chosen goal but not for the counterfactual alternative. The bounding result—that the simplified framework yields an upper bound on optimal choice and a lower bound on conservative choice—carries through unchanged.

Practical Relevance. In the field data, the empirical frequency of cases where the simplified framework could err—scenarios (i) and (iv)—is low. Most employees who choose the high goal satisfy Condition 1, and most who choose a low goal do so despite Condition 1 being satisfied. We therefore expect the bounds to be tight and the main-text estimates of conservative choice to be close to the true values.

Appendix A.2. Convex Effort Costs as a Confound. If some employees misinterpreted the belief elicitation as asking for their likelihood of attaining each goal under levels of effort differing from what they intended to exert (e.g., reporting the probability of attaining a non-selected goal under that goal's optimal effort rather than the effort associated with the chosen goal), then conservative goal choice could

conceivably reflect convex costs of effort rather than attitudes toward risk. We address this concern in the main text by showing that conservative and heterogeneous choice persists when the GoalQuest decision is recast as an explicit choice among economically equivalent financial lotteries with known probabilities, eliminating any role for effort. Here we additionally consider a simple calibration which shows the implausibility of this explanation for explaining the observed patterns.

In the calibration, we modify the risk-neutral subjective EV benchmark to incorporate effort costs expressed as percent reductions in hourly wage. Let σ denote the incremental effort cost of pursuing Goal 2 over Goal 1, and $k \geq 1$ a convexity parameter such that $k\sigma$ is the incremental cost of Goal 3 over Goal 2. For convex effort costs to rationalize a *systematic* preference for lower goals, both σ and k must be non-trivial. Appendix Table A5 reports deterministic hit rates—the share of employees whose observed choice matches the predicted argmax—across a wide grid of values for both parameters.

The calibration illustrates a simple problem. When effort costs are low, the model predicts that most employees should choose Goal 3; when costs are somewhat higher, it predicts that most should choose Goal 1. No common combination of σ and k reproduces the substantial share choosing Goal 2. The model also predicts that employees should choose higher goals when rewards are large relative to the time over which effort costs accumulate. We find no such pattern. In the bottom and top quartiles of Goal 3 reward per program workday, respectively, 29.1 and 29.2 percent choose Goal 1, while 43.6 and 42.4 percent choose Goal 3. With an effort cost equal to only 1 percent of hourly wages, by contrast, the model predicts that Goal 1 choice should fall from 97–100 percent in the bottom quartile to about 15 percent in the top quartile, while Goal 3 choice should rise from 0–3 percent to 63–70 percent, depending on k. Convex effort costs therefore cannot explain either the observed distribution of choices or its stability as relative reward size increases.

Appendix A.3. Structural Robustness to CRRA Utility

Our primary analysis evaluates expected-utility benchmarks under constant absolute risk aversion (CARA). We adopt CARA in the main specifications because it provides a tractable representation of risky choice in the absence of employee-level wealth data. A natural concern, however, is that the weak performance of standard expected utility could reflect the choice of constant absolute, rather than constant relative, risk aversion. In this appendix we address that concern by recharacterizing choice under expected utility assuming constant relative risk aversion (CRRA) utility across a wide range of initial wealth levels and degrees of relative risk aversion, reported in Appendix Table A6.

Specifically, we assume employees evaluate rewards according to CRRA utility of the form

$$u(W+x) = \begin{cases} \dfrac{(W+x)^{1-\rho}}{1-\rho}, & \rho \neq 1, \\ \ln(W+x), & \rho = 1, \end{cases}$$

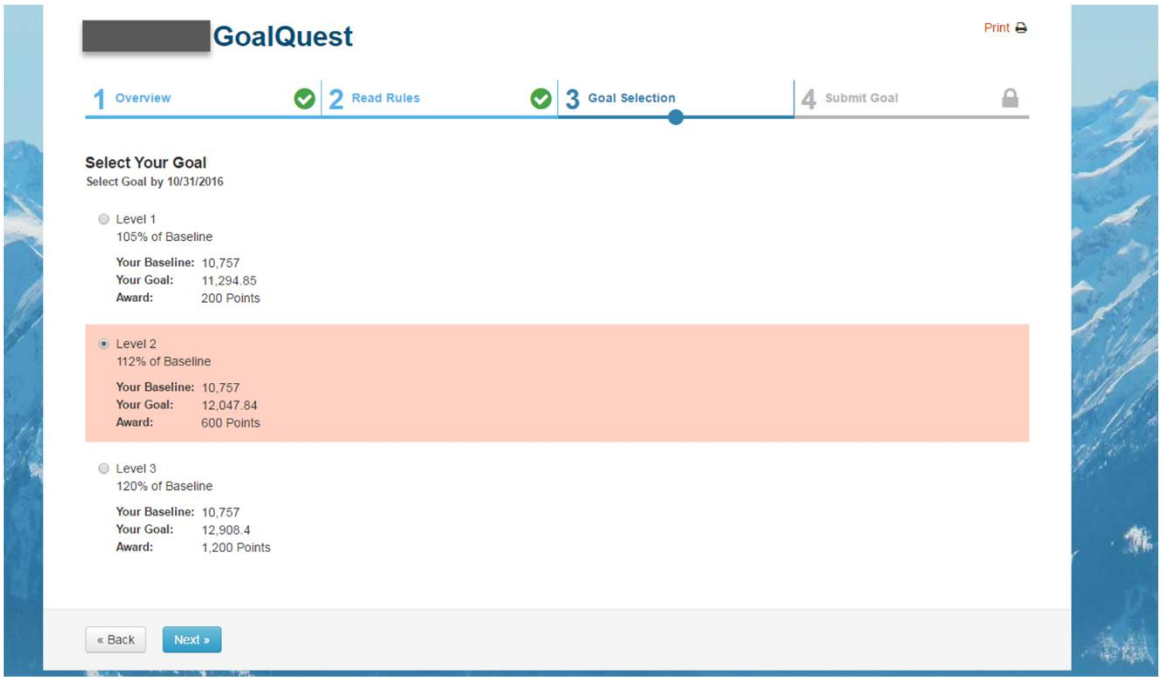
where W denotes initial wealth, x denotes the reward associated with a goal, and ρ denotes the coefficient of relative risk aversion. We consider initial wealth levels ranging from \$1,000 to \$1,000,000 and relative risk aversion values $\rho \in [0.10,50]$. As in the earlier appendix, we summarize the economic meaning of these values using the implied certainty coefficient for a 50/50 bet of $(\$0, \$10{,}000)$ assuming initial wealth of \$25,000. The highlighted region of the table corresponds to the range of relative risk aversion typically viewed as plausible in the literature.

To align this robustness exercise with the main structural analysis, we implement the CRRA benchmarks using the same multinomial logit choice framework used in the paper's horse race. For each combination of beliefs, initial wealth, and relative risk aversion, we compute the expected CRRA utility of each goal and estimate the model's logistic noise parameter by maximum likelihood. We then use the fitted model to generate predicted choice probabilities and report the corresponding deterministic hit rate, defined as the share of observations for which the model assigns the highest probability to the employee's realized choice. We do this separately under rational expectations and subjective expectations.

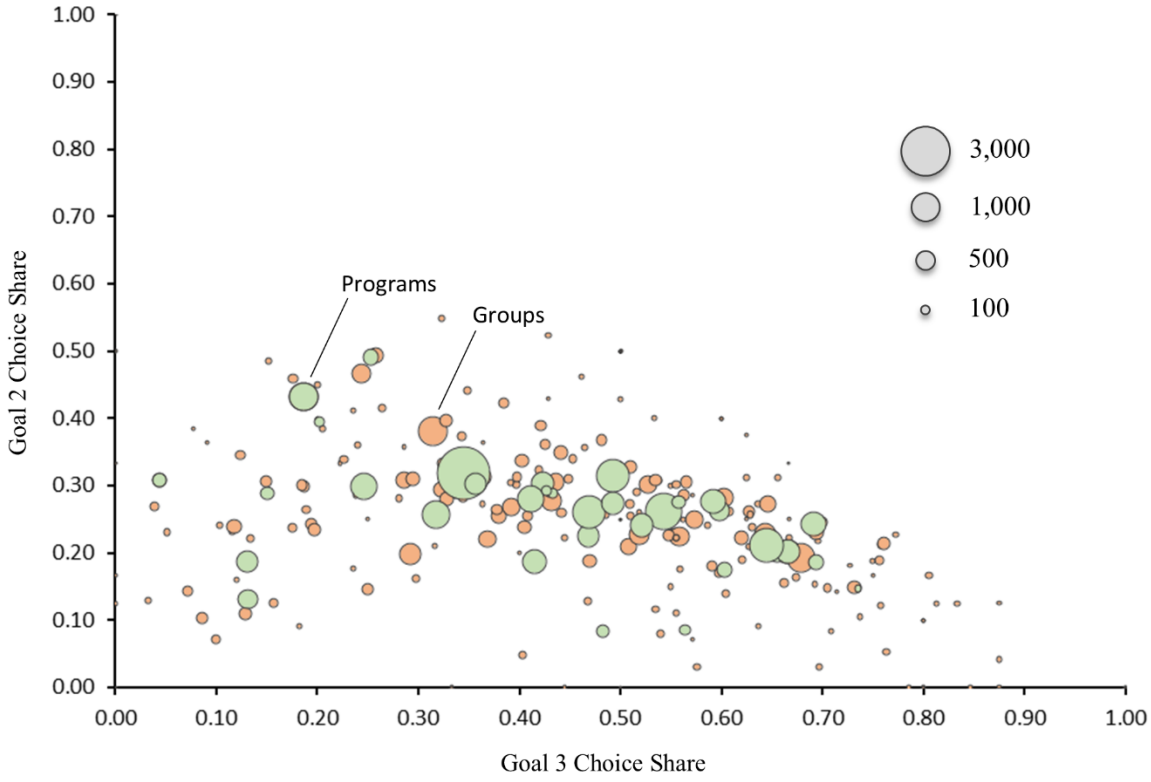
Appendix Table A6 reports the results. Across a wide range of wealth assumptions and plausible values of relative risk aversion, the hit rates are nearly unchanged from the corresponding CARA benchmarks in the main text. Under rational expectations, the CRRA benchmark continues to correctly classify roughly 46 percent of choices over most plausible parameter values. Under subjective expectations, the corresponding hit rate remains close to 50 percent. Only at extreme combinations of very low wealth and very high relative risk aversion do the hit rates move noticeably, and even then the changes are modest.

We interpret these results as showing that the limited descriptive success of expected utility in our setting is not driven by the assumption of CARA rather than CRRA utility. Allowing utility curvature to depend on wealth does not materially improve the benchmark model's ability to account for employee goal choice. Thus, the shortcomings of standard expected utility documented in the main analysis appear to reflect deeper limitations of the benchmark itself rather than the particular utility family used to implement it.

Appendix Figure A1.
Sample Image of GQ Goal Selection Interface



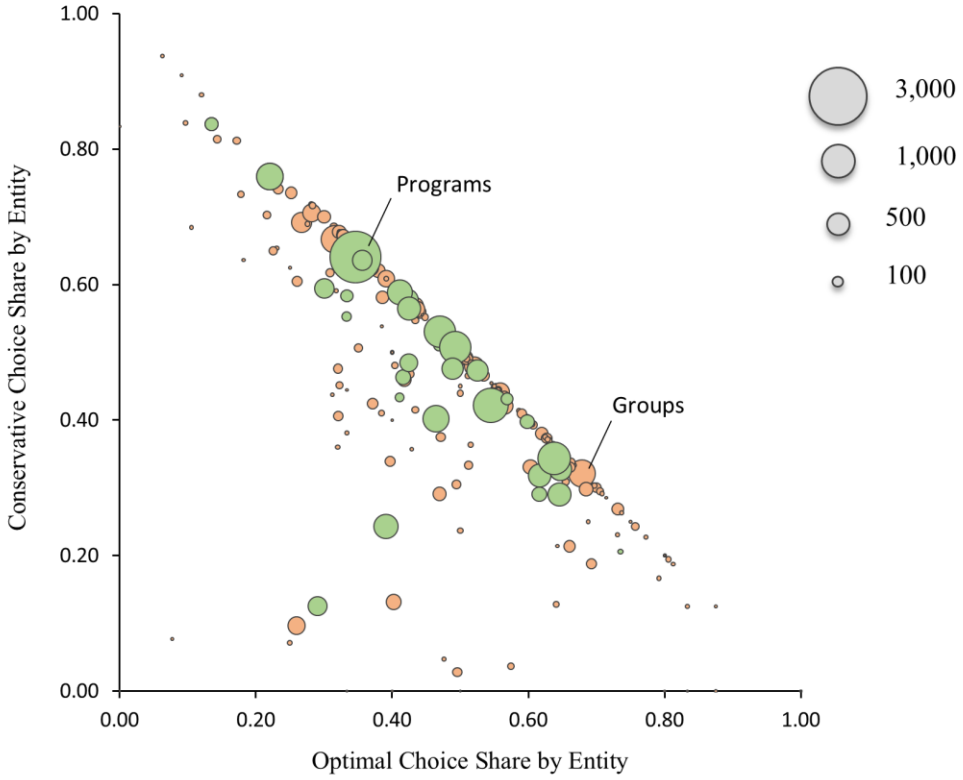
Appendix Figure A2.
Average Program and Group Choice Shares for Goals 2 and 3



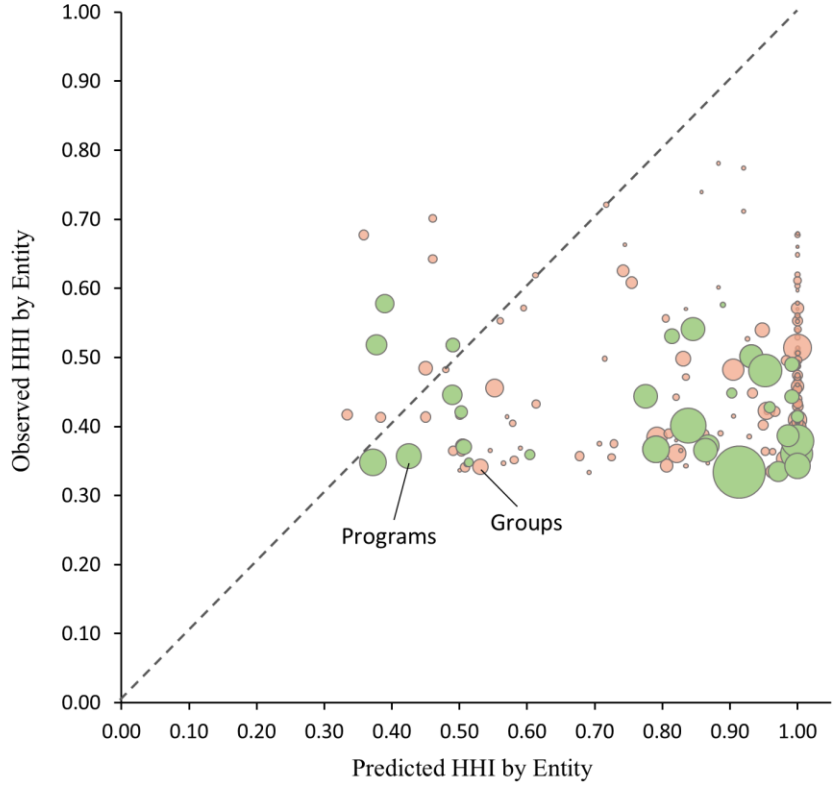
Notes: This figure depicts average choice shares for Goals 2 and 3 for each program (green) and group (orange). Groups with less than 10 employees excluded.

Appendix Figure A3.
Choice Characterization by Program and Group under Expected Value with Rational Expectations

Panel A. Optimal and Conservative Choice Shares



Panel B. Predicted and Observed Choice Heterogeneity (HHI)



Notes: This figure depicts average optimal and conservative choice shares (Panel A) and predicted and observed choice heterogeneity as indicated by HHI (Panel B) under a rational expectations expected value benchmark for each program (green) and group (orange). Groups with less than 10 employees excluded.

Appendix Table A1.

Summary of Sample, Group and Employee Characteristics

		Potential Reward Value	
	All	Below Median	Above Median
Panel A. Sample Overview			
Programs	34	-	-
Groups	232	-	-
Employees	20133	-	-
Firms	18	-	-
Employees per Group (Average)	87	-	-
	(139)		
Employees per Program (Average)	592	-	-
	(587.5)		
Panel B. Group Characteristics (Employee Shares)			
Program Duration			
≤ 30 days	0.39	0.51	0.28
45 to 60 days	0.28	0.12	0.42
≥ 90 days	0.33	0.38	0.29
Potential Reward Value (Estimated $)			
Average	467	150	746
	(482)	(58)	(517)
Median	350	168	525
25th Percentile	175	94	392
75th Percentile	525	175	914
Panel C. Employee Characteristics			
Age [Midpoint of 10-year bins]	36.9	36	37.6
Female	0.46	0.50	0.43
Tenure Category			
< 1 year	0.28	0.32	0.25
1 to 5 years	0.45	0.46	0.43
6 to 10 years	0.14	0.13	0.14
> 10 years	0.13	0.08	0.18
Program-Average Salary (Average) ($1,000s)	70.8	63.2	72.7
Data on Salary Available	0.25	0.10	0.38

Notes: This table summarizes program, group, and employee characteristics for the primary field sample. Panel A reports the number and size of programs and groups. Panel B reports employee-level program duration and potential rewards, where potential reward value is the largest available reward (the Goal 3 reward). Panel C reports demographic characteristics overall and by potential-reward subgroup. Age is imputed from self-reported 10-year bins, gender combines self-reports with first-name inference, and salary is approximated using program-level averages where available.

Appendix Table A2.

Goal Choice, Employee Productivity, and Goal Attainment

		Sample Restricted by Goal Choice		
	All	Goal 1	Goal 2	Goal 3
Panel A. Goal Choice				
Employees	20133	5866	5470	8797
Employee Share	1.00	0.29	0.27	0.44
Potential Reward Value (Average)	466	482	490	442
	(481.5)	(528)	(499)	(434.4)
Panel B. Employee Productivity				
Productivity Relative to Baseline				
Average	1.34	1.12	1.25	1.52
25th Percentile	0.88	0.78	0.89	0.91
50th Percentile	1.01	0.98	1.00	1.04
75th Percentile	1.20	1.11	1.15	1.27
Productivity Relative to Goal 3 Threshold				
Average	0.90	0.68	0.86	1.07
25th Percentile	0.60	0.30	0.63	0.77
50th Percentile	0.89	0.74	0.88	0.95
75th Percentile	1.02	0.95	1.00	1.09
Panel C. Goal Attainment				
Baseline	0.54	0.45	0.53	0.60
Goal 1	0.44	0.32	0.42	0.53
Goal 2	0.36	0.23	0.33	0.47
Goal 3	0.29	0.17	0.25	0.41
Earned Reward (Average)	121	33	92	197
Earned Reward (Average) | Goal Attainment	333	104	277	483

Notes: This table summarizes goal choice, productivity, attainment, and rewards for the primary field sample, overall and by chosen goal. Panel A reports goal choice and potential reward value, defined as the largest available reward (the Goal 3 reward). Panel B reports productivity relative to baseline and to Goal 3; the baseline measure excludes the 18 percent of employees without baseline data. Panel C reports goal attainment and earned rewards.

Appendix Table A3.

Rational-Expectations Benchmark: Construction and Selection Sensitivity

	Observed	Primary Leave-Out	Covariate-Adjusted	Global Choice Cell	Program x Choice Cell	Group x Choice Cell
Panel A. Deterministic RE-EV Choice Implications						
Usable Observations	20,133	20,133	20,133	20,133	20,133	20,066
Probability Coverage	--	100.00%	100.00%	100.00%	100.00%	99.67%
EV Hit Rate	--	0.46	0.45	0.45	0.46	0.52
Conservative Choice Share	--	0.49	0.49	0.55	0.51	0.43
Aggressive Choice Share	--	0.05	0.06	0.00	0.02	0.05
Goal 3 Choice Share	0.44	0.86	0.84	0.99	0.86	0.79
Choice HHI	0.35	0.75	0.72	0.98	0.75	0.64
Employee-Weighted Within-Group Choice HHI	0.42	0.99	0.88	0.99	0.90	0.83
Panel B. Share of RE-Probability Variance Within Group						
Goal 1	--	0.66%	9.60%	--	--	--
Goal 2	--	0.66%	9.65%	--	--	--
Goal 3	--	0.57%	9.29%	--	--	--
Panel C. Selection Stress and Fixed-Performance Reward Accounting						
Share of Primary Conservative Choices Rationalized	--	--	--	0.0%	2.7%	12.4%
Fixed-Performance Reward Sample N	--	3,865	3,852	4,154	3,896	3,557
Foregone Displayed Reward as Share of Optimal Displayed Reward	--	0.45	0.46	0.42	0.45	0.46
Panel D. Subjective Belief Gaps (Percentage Points)						
Goal 1	--	34.18	33.85	--	--	--
Goal 2	--	33.72	33.50	--	--	--
Goal 3	--	32.08	31.99	--	--	--
Panel E. Structural Fits for Ex Ante RE Constructions						
RE-EV Log Likelihood	--	-21812.4	-21817.5	--	--	--
RE-EV sigma	--	419.3	428.0	--	--	--
RE-EU Log Likelihood	--	-21515.3	-21536.4	--	--	--
RE-EU rho	--	0.0032	0.0031	--	--	--
RE-EU sigma	--	67.1	72.5	--	--	--

Notes: This table documents the construction and sensitivity of the rational-expectations expected-value (RE-EV) benchmark. Primary RE probabilities are program-group leave-one-employee-out attainment rates, with covariate-adjusted leave-one-out predictions for two singleton groups; the alternative ex ante specification uses group effects, age, and tenure. Deterministic classifications use each employee's actual reward menu and assign ties to the lower goal. The global, program-by-choice, and group-by-choice columns are employee-favorable post-choice stress tests that condition on the choice being evaluated and therefore are neither forecasting models nor causal estimates. The group-by-choice construction covers 20,066 of 20,133 observations. HHI denotes the Herfindahl-Hirschman Index. The rationalization row gives the share of 9,911 primary conservative choices that become EV-optimal under each stress test. Foregone displayed reward is calculated among employees classified as conservative who attained at least Goal 1, holding realized performance fixed; it is descriptive accounting rather than a causal loss under counterfactual effort. Structural fits are reported only for the two ex ante RE constructions; rho and sigma denote CARA curvature and choice noise.

Appendix Table A4.

Model Fit Across Reward Size and Employee Tenure (Unconstrained)

	RE-EV	EU	LA	Pairwise CN
Reward Value				
Highest Quartile				
Log Likelihood (LL)	-5240	-4654	-4394	-4325
Δ LL Relative to RE-EV	--	586	846	915
Mean Log Likelihood	-1.08	-0.96	-0.91	-0.89
Hit Rate	0.42	0.55	0.60	0.58
Key Parameter	$\sigma = 796$	$\rho = 0.0036$	$\lambda = 0.85$	$\theta = 0.55$
Lowest Quartile				
Log Likelihood (LL)	-5264	-4991	-4721	-4629
Δ LL Relative to RE-EV	--	273	543	635
Mean Log Likelihood	-1.09	-1.04	-0.98	-0.96
Hit Rate	0.40	0.48	0.54	0.56
Key Parameter	$\sigma = 91$	$\rho = 0.033$	$\lambda = 4.26$	$\theta = 0.70$
Employee Tenure				
10+ Years				
Log Likelihood (LL)	-2916	-2623	-2558	-2497
Δ LL Relative to RE-EV	--	293	358	419
Mean Log Likelihood	-1.09	-0.98	-0.96	-0.94
Hit Rate	0.40	0.53	0.59	0.57
Key Parameter	$\sigma = 981$	$\rho = 0.0056$	$\lambda = 0.75$	$\theta = 0.40$
< 1 Year				
Log Likelihood (LL)	-6177	-5876	-5585	-5348
Δ LL Relative to RE-EV	--	301	592	829
Mean Log Likelihood	-1.08	-1.03	-0.98	-0.94
Hit Rate	0.45	0.49	0.57	0.58
Key Parameter	$\sigma = 319$	$\rho = 0.0032$	$\lambda = 1.08$	$\theta = 0.75$

Notes: This table reports subgroup-specific, unconstrained model fits by potential reward value and tenure. Reward subgroups are the highest and lowest quartiles of the largest available reward; tenure subgroups are more than 10 years and less than 1 year. Each model is estimated separately within each subgroup. Reported statistics are subgroup log likelihood, mean log probability assigned to the observed choice, deterministic hit rate, and the key parameter estimate. RE-EV uses program-group leave-one-employee-out attainment rates; EU is subjective CARA expected utility; LA is loss aversion; and Pairwise CN is pairwise contingency neglect. The parameters rho, lambda, and theta denote absolute risk aversion, loss aversion, and contingency neglect, respectively.

Appendix Table A5.

Optimal Goal Choice Shares under Subjective EV Benchmark with Effort Costs

	Baseline Effort Cost Increment as % of Wage						
Convexity (k)	0%	1%	3%	5%	10%	25%	50%
1.00	0.50	0.50	0.33	0.30	0.29	0.29	0.29
1.10	0.50	0.50	0.32	0.30	0.29	0.29	0.29
1.25	0.50	0.48	0.32	0.30	0.29	0.29	0.29
1.50	0.50	0.47	0.32	0.30	0.29	0.29	0.29
2.00	0.50	0.42	0.31	0.29	0.29	0.29	0.29
5.00	0.50	0.36	0.30	0.29	0.29	0.29	0.29

Notes: This table reports the share of observed choices that are optimal under a subjective expected-value benchmark incorporating effort costs. Baseline effort-cost increment is the increase in hourly effort cost for Goal 2 relative to Goal 1, expressed as a percentage of wage. The convexity parameter k scales the additional effort cost of Goal 3 relative to the Goal 2 increment. Calculations use employees' elicited beliefs and assume a wage of $25 per hour and eight working hours per day. For example, a 25-working-day program with a 10 percent baseline increment and k = 1.5 implies total effort costs of $0, $500, and $1,250 for Goals 1, 2, and 3.

Appendix Table A6.

Goal Choice Hit Rates under Expected CRRA Utility Benchmarks

		Rational Expectations - Initial Lifetime Wealth						
ρ	CC(0/10k)	$1,000	$10,000	$25,000	$50,000	$100,000	$500,000	$1,000,000
0.10	0.99	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.25	0.98	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.50	0.96	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.75	0.94	0.45	0.45	0.45	0.45	0.45	0.45	0.45
1.00	0.92	0.45	0.45	0.45	0.45	0.45	0.45	0.45
1.50	0.87	0.45	0.45	0.45	0.45	0.45	0.45	0.45
2.50	0.79	0.46	0.45	0.45	0.45	0.45	0.45	0.45
5.00	0.61	0.46	0.45	0.45	0.45	0.45	0.45	0.45
10.00	0.37	0.42	0.46	0.45	0.45	0.45	0.45	0.45
50.00	0.07	0.30	0.45	0.45	0.46	0.45	0.45	0.45

		Subjective Expectations - Initial Lifetime Wealth						
ρ	CC(0/10k)	$1,000	$10,000	$25,000	$50,000	$100,000	$500,000	$1,000,000
0.10	0.99	0.50	0.50	0.50	0.50	0.50	0.50	0.50
0.25	0.98	0.50	0.50	0.50	0.50	0.50	0.50	0.50
0.50	0.96	0.50	0.50	0.50	0.50	0.50	0.50	0.50
0.75	0.94	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1.00	0.92	0.51	0.50	0.50	0.50	0.50	0.50	0.50
1.50	0.87	0.51	0.50	0.50	0.50	0.50	0.50	0.50
2.50	0.79	0.52	0.50	0.50	0.50	0.50	0.50	0.50
5.00	0.61	0.53	0.50	0.50	0.50	0.50	0.50	0.50
10.00	0.37	0.53	0.51	0.50	0.50	0.50	0.50	0.50
50.00	0.07	0.48	0.53	0.52	0.51	0.50	0.50	0.50

Notes: This table reports deterministic hit rates for expected-utility benchmarks with CRRA utility across assumptions about initial wealth and relative risk aversion. Each cell is the share of observed choices equal to the deterministic optimum for the stated beliefs, wealth, and curvature; the likelihood model's logit noise parameter is estimated separately for each specification. The second column reports the certainty coefficient for a 50/50 bet paying $0 or $10,000 at $25,000 of initial wealth, defined as the certainty equivalent divided by expected value. The highlighted columns mark the range of relative risk aversion emphasized in the analysis. Panel A uses covariate-adjusted program-group leave-one-employee-out probabilities, and Panel B uses elicited subjective beliefs.

Appendix Table A7.

Optimal Goal Choice Shares for Gain-Loss Utility Benchmarks by Candidate Reference Point

	Gain-Loss Utility ($\alpha = 0.88$; $\eta = 0$)			Consumpton + Gain-Loss Utility ($\lambda = 2.25$)		
Candidate Reference Points	$\lambda = 1.50$	$\lambda = 2.25$	$\lambda = 3.00$	$\eta = 1$	$\eta = 3$	$\eta = 5$
Panel A. Prospect Independent						
Status Quo (0)	0.50	0.50	0.50	0.50	0.50	0.50
High Probability (Goal 1)	0.52	0.54	0.55	0.51	0.50	0.50
Compromise Goal (Goal 2)	0.50	0.52	0.52	0.50	0.50	0.50
Maximum Reward (Goal 3)	0.49	0.49	0.49	0.49	0.49	0.49
Maximum High Certainty	0.51	0.51	0.51	0.50	0.50	0.50
Panel B. Prospect-Dependent						
Reward of Chosen Goal	0.29	0.29	0.29	0.59	0.56	0.54
Expected Value of Chosen Goal	0.40	0.26	0.26	0.54	0.50	0.50
Reward of Chosen Goal + 1	0.55	0.55	0.55	0.53	0.53	0.52
Reward of Chosen Goal - 1	0.46	0.43	0.42	0.58	0.54	0.53
Regret (Expected Max Counterfactual)	0.50	0.50	0.50	0.50	0.50	0.50

Notes: This table reports optimal-choice shares for gain-loss utility benchmarks across candidate reference points, functional forms, and parameter values, using subjective attainment beliefs. The first block uses prospect-theory gain-loss utility and varies the loss-aversion parameter lambda. The second block uses composite utility, an additive combination of consumption and gain-loss utility, and varies the consumption-utility scaling factor n; n = 0 implies gain-loss utility alone. Panel A considers prospect-independent reference points, and Panel B considers prospect-dependent reference points. Each cell reports the share of observed choices that the specified benchmark classifies as optimal.

Appendix Table A8.

Joint Estimates of Pairwise Contingency Neglect and Alternative Mechanisms

	Loss Aversion		Common 0.50 Anchor		Generic Upgrade Penalty	
	$\rho = 0.001$	ρ Estimated	$\rho = 0.001$	ρ Estimated	Constant ($\rho = 0.001$)	Proportional ($\rho = 0.001$)
Joint-Model Parameters						
Added Parameter Estimate	$\lambda = 1.224$	$\lambda = 1.000$	$\kappa = 0.000$	$\kappa = 0.000$	$\delta = 0.000$	$\eta = 0.0165$
One-Sided 95% Profile Upper Bound	--	--	0.00382	0.000744	0.0924	0.0295
Number of Estimated Parameters	3	4	3	4	3	3
Fit Relative to PCN Alone						
Δ Log Likelihood	+25.971	0.000	0.000	0.000	0.000	+1.957
BIC	38,338.32	37,363.82	38,390.26	37,363.82	38,390.26	38,386.35
Δ BIC	-42.03	+9.91	+9.91	+9.91	+9.91	+6.00
Boundary and Held-Out Diagnostics						
Zero-Boundary Specifications	$\kappa = \delta = 0$ reproduces PCN's likelihood; each extra parameter raises BIC by ln(20,133) = 9.91.					
Loss Aversion with Estimated Curvature	$\lambda = 1$ in all 34 leave-one-program-out fits; Δ LOPO log score = 0.000.					

Notes: This table jointly estimates Pairwise CN with loss aversion, compression toward a common 0.50 anchor, or a generic upgrade penalty, using 20,133 employee choices, subjective beliefs, and a common ascending sequential-logit architecture. Delta log likelihood and Delta BIC compare each joint model with Pairwise CN alone under the same treatment of curvature; positive Delta log likelihood and negative Delta BIC favor the joint model. Fixed-curvature specifications impose rho = 0.001; otherwise rho is estimated on [0, 0.005]. Pairwise CN alone has BIC 38,380.35 with fixed curvature and 37,353.91 with estimated curvature. The added parameters are lambda >= 1 for loss aversion, a nonnegative anchor weight kappa, a nonnegative constant penalty delta, and a nonnegative proportional penalty eta. A zero anchor or penalty reproduces Pairwise CN and raises BIC by ln(20,133) = 9.91 for the redundant parameter. One-sided profile bounds are reported for the anchor and penalty parameters; loss-aversion profile bounds are not reported.

Appendix Table A9.

Structural Model Horserace - Alternative Pairwise Heuristics under a Common Sequential Stopping Rule (Constrained Curvature)

	Pairwise EU	Compromise Effect	Positional Bias	Salience	Pairwise CN
Model Fit Statistics					
Log Likelihood (LL)	-19999	-19866	-19866	-19999	-19180
AIC	40003	39738	39740	40005	38367
BIC	40018	39762	39771	40028	38390
ΔLL Relative to RE-EV	1813	1946	1947	1813	2632
ΔLL Relative to Pairwise EU	--	133	133	0	819
Hit Rate	0.52	0.53	0.53	0.52	0.56
Predicted Goal 3 Choice Share (Observed: 0.44)	0.53	0.57	0.56	0.53	0.44
Residual Conservative Choice Share	0.24	0.27	0.27	0.24	0.18
Share of RE-EV Gap Closed					
Conservative Choice	0.88	0.78	0.80	0.88	1.03
Herfindahl-Hirschman Index	0.76	0.76	0.78	0.76	0.88
Key Parameters	$\rho = 0.001$	$d = 62.97$, $\rho = 0.001$	$\beta_{12} = 14.40$, $\beta_{23} = -16.89$, $\rho = 0.001$	$\psi \approx 0$, $\rho = 0.001$	$\theta = 0.812$, $\rho = 0.001$

Notes: This table compares alternative pairwise heuristics in the primary field sample under a common ascending sequential-logit rule. Pairwise EU is the sequential CARA expected-utility benchmark; Compromise Effect adds a middle-option bonus; Positional Bias adds identified transition-specific shifts; Salience adds a sequential salience distortion; and Pairwise CN adds contingency neglect. In every column, rho is estimated on [0, 0.001] and the upper bound binds; AIC and BIC count rho, choice noise, and all behavioral parameters. Positional Bias is reported using identified transition shifts because its original scale representation is redundant. Reported outcomes are log likelihood, AIC, BIC, deterministic hit rate, predicted Goal 3 share, residual conservative-choice share, and closure of the RE-EV-to-data gaps in conservative choice and HHI. Delta log likelihood is reported relative to both unconstrained RE-EV and the first-column sequential Pairwise EU comparator. Gap closure equals 1 at an exact match and 0 at RE-EV, and can exceed 1 with overshooting.

Appendix Table A10.

Structural Model Horserace - Low-Curvature Calibration

	Rational Expectations		Subjective Expectations				
	EV	EU	EV	EU	RD-EU	LA	Pairwise CN
Model Fit Statistics							
Log Likelihood (LL)	-21812	-21630	-20935	-20275	-20275	-19666	-19188
AIC	43627	43264	41872	40554	40556	39338	38382
BIC	43635	43280	41880	40569	40579	39362	38406
ΔLL Relative to RE-EV	--	182	877	1538	1538	2146	2625
ΔLL Relative to Subjective EU	-1538	-1355	-660	--	0	609	1087
Hit Rate	0.46	0.46	0.50	0.51	0.51	0.59	0.55
Predicted Goal 3 Choice Share (Observed: 0.44)	0.86	0.78	0.87	0.84	0.84	0.55	0.45
Residual Conservative Choice Share	0.49	0.46	0.48	0.47	0.47	0.24	0.19
Share of RE-EV Gap Closed (1.00 = exact match)							
Conservative Choice	0.00	0.17	0.21	0.26	0.26	0.82	1.02
Herfindahl-Hirschman Index	0.00	0.28	-0.03	0.07	0.07	0.86	0.87
Key Parameters	s = 419	ρ = 0.001	s = 315	ρ = 0.001	α = 1.00 ρ = 0.001	λ = 1.00 α = 1.00	theta = 0.75 rho = 0.001 sigma = 54.22

Notes: This table reports a low-curvature calibration of field models under rational and subjective expectations. EV denotes expected value; EU denotes CARA expected utility; RD-EU uses Prelec probability weighting; LA denotes loss aversion; Heterogeneous EU is a three-type latent-class model; and Pairwise CN denotes pairwise contingency neglect. Reported outcomes are log likelihood, AIC, BIC, deterministic hit rate, predicted Goal 3 share, residual conservative-choice share, and closure of the RE-EV-to-data gaps in conservative choice and HHI. Residual conservative choice means choosing below the model's deterministic prediction; gap closure equals 1 at an exact match and 0 at RE-EV, and can exceed 1 with overshooting. RE-EV uses program-group leave-one-employee-out attainment rates with covariate-adjusted predictions for two singleton groups. Pairwise CN searches theta over {0, 0.25, 0.50, 0.75, 1} with rho capped at 0.001 and selects theta = 0.75, log likelihood = -19,187.902, and sigma = 54.224. This is a coarse-grid calibration rather than unrestricted estimation or a continuous mechanism decomposition.

Appendix Table A11.

Structural Comparison between Primary and Expansive Samples under Rational Expectations

	Primary Sample		Expansive Sample		Difference	
	RE-EV	RE-EU	RE-EV	RE-EU	RE-EV	RE-EU
Mean Probability of Observed Choice	0.34	0.35	0.34	0.34	0.00	-0.01
Hit Rate	0.46	0.45	0.42	0.44	-0.03	-0.01
Predicted Goal 3 Choice Share (Observed: 0.44)	0.86	0.73	0.87	0.74	0.01	0.01
Residual Conservative Choice Share	0.49	0.44	0.54	0.48	0.04	0.04
Share of RE-EV Gap Closed						
Conservative Choice	0.00	0.29	0.00	0.29	0.00	-0.01
Herfindahl-Hirschman Index	0.00	0.45	0.00	0.43	0.00	-0.02
Key Parameters	$\sigma = 419$	$\rho = 0.003$	$\sigma = 596$	$\rho = 0.002$	--	--

Notes: This table compares the primary and expansive field samples under rational-expectations benchmarks. RE-EV denotes expected value, and RE-EU denotes CARA expected utility. Reported outcomes are mean probability assigned to the observed choice, deterministic hit rate, predicted Goal 3 share, residual conservative-choice share, and closure of the RE-EV-to-data gaps in conservative choice and the Herfindahl-Hirschman Index (HHI). Residual conservative choice means choosing below the model's deterministic prediction; gap closure equals 1 at an exact match and 0 at RE-EV, and can exceed 1 with overshooting. Primary-sample probabilities use program-group leave-one-employee-out rates with covariate-adjusted predictions for two singleton groups; expansive-sample estimates use the corresponding archived sensitivity specification.

Appendix Table A12.

Robustness of Structural Models to Counterfactual Effort

Model	Holdout Hit Rates			Holdout Mean Log Score	
	Baseline (No Effort)	Median across Effort Grid	Maximum across Effort Grid	Baseline (No effort)	Training
EV–RE	0.46	0.30	0.46	-1.08	-1.08
EU–RE	0.45	0.30	0.45	-1.07	-1.07
EV–SB	0.50	0.31	0.50	-1.04	-1.06
EU–SB	0.51	0.31	0.51	-1.01	-1.11
Rank-dependent EU (Prelec)	0.51	0.31	0.51	-1.01	-1.11
Loss aversion	0.58	0.30	0.58	-0.98	-0.98

Notes: This table tests the sensitivity of structural-model predictions to counterfactual effort adjustments. Reported beliefs are treated as conditional on the effort associated with each employee's observed goal choice. The 65 prespecified scenarios vary both the effect of effort on attainment beliefs and the monetary-equivalent cost of pursuing higher goals. No effort adjustment uses reported beliefs without modification and sets effort costs to zero. Each model is re-estimated on the training sample and evaluated on the holdout sample. Median and maximum hit rates summarize the effort grid; the maximum is a descriptive sensitivity bound rather than a model-selection result. Training-selected effort reports holdout performance for the scenario selected by training-sample likelihood. Mean log score is the average log probability assigned to the observed choice, with higher values indicating better predictive performance.

Appendix Table A13.

Experiment Overview

Experiment	Date	Sample	Design and Treatments	Incentives (beyond base pay)	Analytic N	Primary Role
Experiment A	July 2026	Prolific; employed U.S. adults, 25 to 55 years	Between-subjects choice from one three-option nested menu: hypothetical GQ, hypothetical RQ, incentivized RQ, or incentivized RQ with accurate pairwise conditional probabilities	Yes; RQ-I and RQ-IP offered consequential bonus payments	1140	[Pre-registered] Validates the lottery representation; measures conditional beliefs and comparison processes; provides an incentivized test of accurate conditional information
Experiment B	May 2019	Amazon M-Turk; employed U.S. adults, 25 to 55 years	Within-subject choices from six three- or four-option GoalQuest-style effort menus varying difficulty, reward structure, or menu size	Yes; one menu selected randomly for bonus payment	277 / 1,662 choices	Tests whether PCN and stable preference-based mechanisms organize repeated choices across menus
Experiment C	July 2022	Amazon M-Turk; employed U.S. adults, 25 to 55 years	Two modules: belief and process elicitation; and a three-arm likelihood-presentation experiment using absolute, accurate conditional, or biased conditional probabilities	No	815	Directly measures conditional-belief errors and tests whether accurate conditional information improves choice
Experiment D	August 2023	Amazon M-Turk; employed U.S. adults, 25 to 55 years	Between-subjects hypothetical prescription-drug plan choice: informationally equivalent Baseline (marginal probabilities) and Accurate-Conditionals frames, plus a deliberately diagnostic PCN-Implied-Conditionals arm	No; plan choice was hypothetical	432	Tests whether directly supplying the comparison-relevant conditional changes use or weighting of a commonly identified tail state; does not identify market underinsurance or welfare
Experiment E	October 2022	Amazon M-Turk; employed U.S. adults, 25 to 55 years	Between-subjects hypothetical homeowners-insurance choice under four decision-equivalent, not informationally equivalent frames that display different probability components while preserving the benchmark-optimal Basic Plan over the maintained belief/preference domain: no information, full nested, no-damage only, or severity conditional only	No	302	Tests whether decision-equivalent probability displays shift coverage choice in the opposing directions implied by outer- versus inner-node focality; identifies frame effects, not latent mechanism or welfare

Notes: This table summarizes the samples, designs, incentives, and primary evidentiary roles of Experiments A-E. Analytic N refers to participants remaining after the applicable exclusions; Experiment B also reports the number of repeated choices. Experiment A applies the preregistered exclusions for potential bots and excessively fast responses. Experiment B excludes incomplete cases and observations with inconsistent beliefs. Experiment C includes the 815 participants who passed all screens, including the comprehension check; treatment randomization occurred after that check. Experiment D retains completed final-design responses with valid completion identifiers and the stored attention pass. Experiment E retains 302 responses with nonmissing completion identifiers across four decision-equivalent probability frames. GQ denotes GoalQuest, and RQ denotes RewardQuest.

Appendix Table A14.

Experiment B: Repeated-Choice Rationalization Across Models (Descriptive)

Panel A. In-Sample Rationalization Across Models

			Loss Aversion			Contextual Sorting Heuristics	
	EU-SB	RD-EU	Estimated lambda	Personal lambda	Pairwise CN	Ability	Taste for Competition
All Choices (6/6)	0.14	0.16	0.29	0.16	0.31	0.07	0.06
Nearly All Choices (5+/6)	0.29	0.35	0.43	0.25	0.57	0.09	0.09
Most Choices (4+/6)	0.43	0.47	0.58	0.39	0.72	0.27	0.28

Panel B. Overlap between Pairwise CN and Loss Aversion

Threshold	Pairwise CN Alone	Loss Aversion Alone	Both Models	Neither Model
All Choices (6/6)	0.17	0.15	0.14	0.54
Nearly All Choices (5+/6)	0.27	0.13	0.30	0.30

Notes: This table reports descriptive in-sample rationalization of Experiment B participants' six repeated choices. Panel A gives the share rationalized by each model under the indicated thresholds. EU-SB denotes subjective expected utility; RD-EU denotes rank-dependent expected utility; Estimated Loss Aversion fits a participant-specific loss-aversion coefficient; Personal Loss Aversion uses the independently elicited coefficient; and Pairwise CN fits a participant-specific contingency-neglect parameter. Panel B classifies participants according to whether Pairwise CN, estimated loss aversion, both, or neither rationalizes their choices at each threshold. Because the same choices are used for fitting and evaluation and the models differ in flexibility, these comparisons are descriptive rather than predictive.

Appendix Table A15.

Experiment C: Conditional-Belief Diagnostics and Direct-Rival Tests

Panel A. Reported Pairwise Conditionals and Bayesian Benchmarks

Measure	Goal 2 | Goal 1 Attained	Goal 3 | Goal 2 Attained
Reported Conditional	63.0%	59.0%
Bayesian Benchmark	81.5%	76.5%
Difference	-18.5%	-17.5%
N	398	397
p-value	<0.001	<0.001

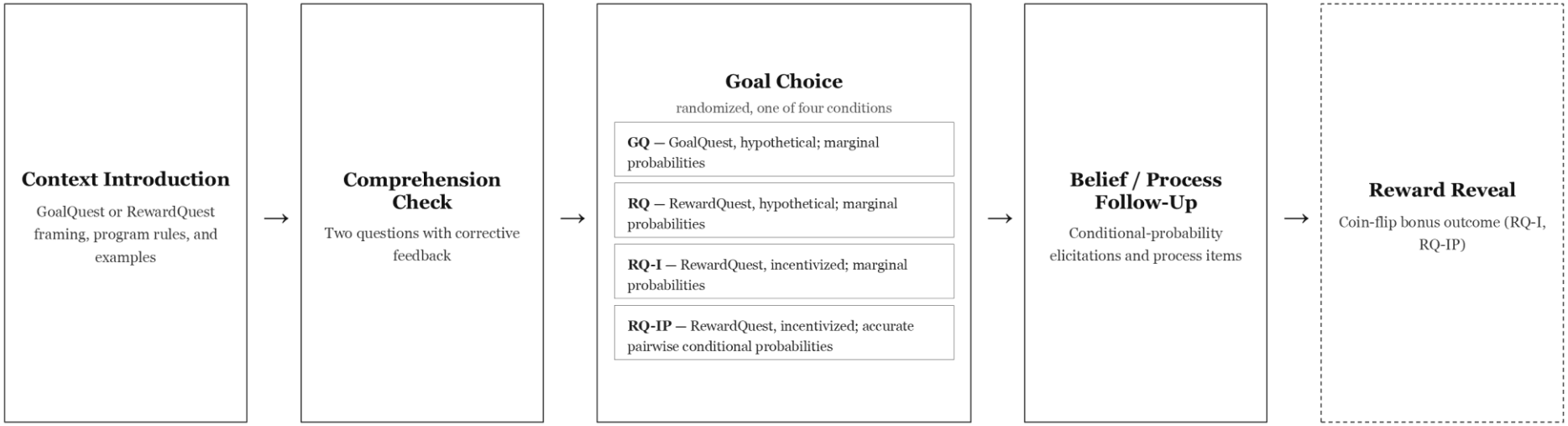
Panel B. Marginal Anchor versus Generic 0.50 Compression

Metric	0.50 Compression	PCN Marginal Anchor	Joint PCN + 0.50	Unrestricted Diagnostic
Behavioral / Incremental Estimate	$\kappa = 0.501$	$\theta = 0.816$	$\theta = 0.567$; $\kappa = 0.261$	Marginal = 0.526
Respondent-Clustered SE	0.026	0.040	(0.048, 0.027)	0.053
In-Sample RMSE	0.2107	0.1979	0.1854	0.1851
Mean Held-Out RMSE	0.2111	0.1984	0.1861	0.1861
Mean Held-Out MAE	0.1672	0.1405	0.1402	0.1404
Key Diagnostic		< 0.50 RMSE, all 20 splits	$p = 0.151$	$p < 0.001$

Notes: This table reports conditional-belief diagnostics and direct-rival tests from Experiment C. Panel A compares reported adjacent-goal conditionals with participant-specific Bayesian conditionals implied by respondents' marginal beliefs; Difference is reported minus Bayesian, and p-values test a zero mean difference. Panel B uses 795 defined conditional reports from 398 participants. Cross-validation consists of 20 repetitions of respondent-grouped 10-fold splits. The 0.50 Compression model shrinks reports toward one-half, the PCN Marginal Anchor model shrinks them toward the corresponding marginal belief, and the joint model permits both components. The one-parameter marginal-anchor model has lower held-out RMSE than equally parsimonious 0.50 compression in all 20 splits. The joint model improves fit further, indicating that the marginal anchor is an important predictive component rather than a complete one-parameter description of every report. Standard errors are clustered by respondent.

Experimental Appendix A.1 — Schematic Overview

Experiment A: Goal-choice validation with four experimental conditions (July 2026 · Prolific · analytic N = 1,140 · hypothetical in GQ and RQ, incentivized in RQ-I and RQ-IP)



Notes: Omitted: consent, Prolific ID collection.

Experimental Appendix A.2 — Introduction Screens

(a) GoalQuest arm (GQ)

SCREEN 1 · CONTEXT INTRODUCTION

Thanks for agreeing to take this survey. For the next series of questions, please imagine that you're a sales employee at ABC, a large US consumer goods company. To encourage greater sales, ABC is set to launch a rewards program called **GoalQuest**. GoalQuest is a proprietary employee-rewards program used by companies around the world. Next, we will tell you how the program works.

SCREEN 2 · PROGRAM RULES

In **GoalQuest**, you will choose one sales goal from a menu of three options. The goals differ in both reward size and difficulty: higher goals pay larger rewards, but require selling more units and are therefore harder to achieve. For each goal, we will show you the chance that you would sell enough units to achieve that goal.

At the end of the sales period, your total number of units sold will determine whether you earn your selected reward. If you sell at least as many units as the goal you selected, you earn the reward for that goal. If you sell fewer units than the goal you selected, you earn nothing.

(b) RewardQuest arms (RQ, RQ-I, RQ-IP)

SCREEN 1 · CONTEXT INTRODUCTION

Thanks for agreeing to take this survey. For the next series of questions, please imagine that you're a sales employee at ABC, a large US consumer goods company. To encourage greater sales, ABC is set to launch a rewards program called **RewardQuest**. RewardQuest is similar to a proprietary employee-rewards program used by companies around the world. Next, we will tell you how the program works.

SCREEN 2 · PROGRAM RULES

In **RewardQuest**, you will choose one reward from a menu of three options. The rewards differ in both size and likelihood of being earned: larger rewards require more RewardQuest points and are therefore less likely to be earned.

After you choose a reward, a number from 1 to 100 will be drawn at random. This randomly drawn number is the only thing that determines your RewardQuest points. If your points are at least as high as the number of points needed for the reward you selected, you earn that reward. If your points are below the number of points needed, you earn nothing.

Notes: Omitted: nested-structure screen, three worked examples, comprehension questions. All three RewardQuest arms share the same introduction.

Experimental Appendix A.3 — Choice Interventions

(a) GoalQuest condition (GQ)

Now we'd like you to select a GoalQuest goal from a real-life menu. To help you make your decision, we will tell you your chances of achieving each goal. Please assume these probabilities are accurate.

Which GoalQuest goal would you choose from the menu below? Please answer as if you were an actual employee participating in this program.

- **GOAL 1**
  Sales Goal: Sell **105 units or more**
  Reward: **$150**
  Your chances of achieving Goal 1: 83%
- **GOAL 2**
  Sales Goal: Sell **110 units or more**
  Reward: **$450**
  Your chances of achieving Goal 2: 74%
- **GOAL 3**
  Sales Goal: Sell **115 units or more**
  Reward: **$900**
  Your chances of achieving Goal 3: 65%

(b) RewardQuest condition (RQ)

Now we'd like you to select a reward from a real-life RewardQuest menu. To help you make your decision, we will tell you your chances of earning each reward. Please assume these probabilities are accurate.

Which RewardQuest reward would you choose from the menu below? Please answer as if you were an actual employee participating in this program.

- **REWARD 1**
  Points needed: **18 or more**
  Reward: **$150**
  Your chances of earning Reward 1: 83%
- **REWARD 2**
  Points needed: **27 or more**
  Reward: **$450**
  Your chances of earning Reward 2: 74%
- **REWARD 3**
  Points needed: **36 or more**
  Reward: **$900**
  Your chances of earning Reward 3: 65%

(c) Incentivized RewardQuest (RQ-I)

RQ-I · REWARDQUEST, INCENTIVIZED

As in (b), with the introduction adding: **"To make this more realistic, at the end of the study, Qualtrics will flip a coin and if it lands on heads we will give you a bonus equal to 1/2 cent for every dollar of displayed reward you earn."**

(d) Incentivized RewardQuest w/ Accurate Conditionals (RQ-IP)

RQ-IP · INCENTIVIZED, ACCURATE PAIRWISE CONDITIONALS

As in RQ-I, but Rewards 2 and 3 instead display: **"If you earn Reward 1, your chances of earning Reward 2 are: 89%"** and **"If you earn Reward 2, your chances of earning Reward 3 are: 88%."**

Notes: Each participant saw exactly one of the four screens; menus are identical in reward amounts and (implied) probabilities, and the arms vary framing, incentives, and probability displays.

Experimental Appendix A.4 — Conditional-Belief Elicitations

(a) Reward 2 given Reward 1

CONDITIONAL BELIEF · R2 | R1

Suppose your RewardQuest points are high enough to earn Reward 1. What are the chances that your points are also high enough to earn Reward 2? Please enter a percent chance from 1 to 100.

(b) Reward 3 given Reward 2

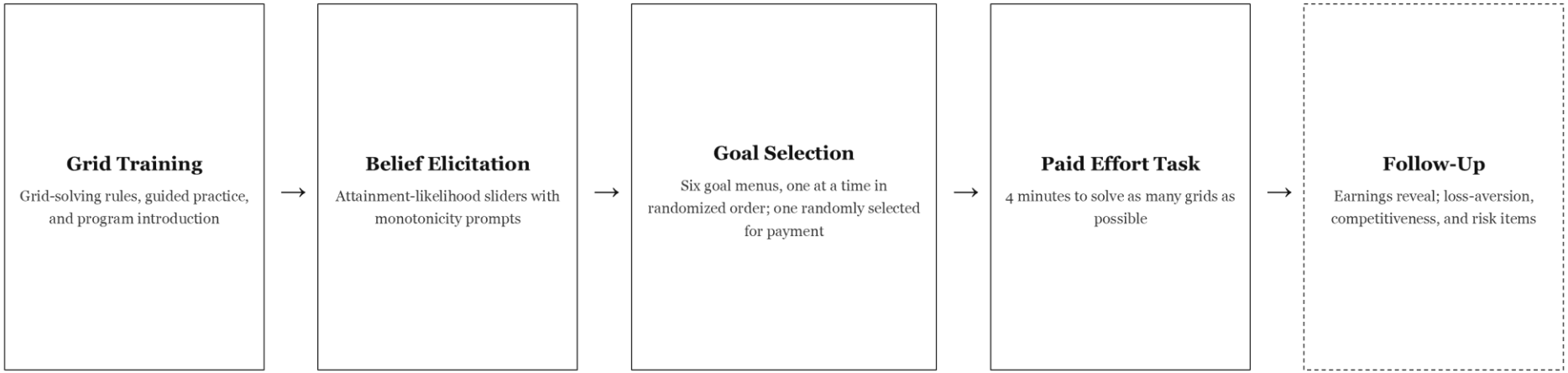
CONDITIONAL BELIEF · R3 | R2

Finally, suppose your RewardQuest points are high enough to earn Reward 2. What are the chances that your points are also high enough to earn Reward 3? Please enter a percent chance from 1 to 100.

Notes: Omitted: open-ended decision description, decision-process checklist. Asked in every arm immediately after choice, each on its own page; the GoalQuest arm substitutes "Goal" for "Reward" and refers to units sold.

Experimental Appendix B.1 — Schematic Overview

Experiment B: Repeated incentivized goal choice in a real-effort grid task (May 2019 · Amazon Mechanical Turk · analytic N = 277, 1,662 choices · incentivized)



Notes: Omitted: consent, demographics. Participants were employed U.S. adults.

Experimental Appendix B.2 — Instructions

(a) Grid rules and example

SCREEN 1 · GRID RULES AND EXAMPLE

In this study, you can earn bonus payments by solving simple puzzles, called **"grids."** Every grid you will see contains several numbers ranging between 0.00 and 10.00. To solve a grid you must find the unique pair of numbers within the grid that sum to exactly **10.00**. For example, in the grid below, the correct numbers are **8.98** and **1.02**.

4.51	8.98	2.30
6.77	1.02	9.16
3.05	5.44	7.83

(b) GoalQuest introduction and example menu

SCREEN 2 · GOALQUEST OVERVIEW

Welcome to **GoalQuest!** You will be given 4 minutes to correctly solve as many puzzles as you can, earning a base pay of **$0.02** per solved grid. As part of the GoalQuest program, you can also earn a **bonus**: you will select one of three goal levels, like the example menu below, and if you meet your goal you earn the bonus for that goal level in addition to your base pay.

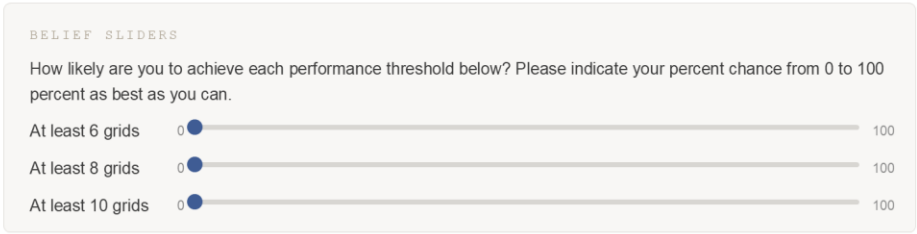
Goal Level:	Level 1	Level 2	Level 3
Performance Threshold:	6 grids	8 grids	10 grids
Bonus:	$0.05	$0.15	$0.30

If you do not meet your goal, you will not receive any bonus in addition to the base pay.

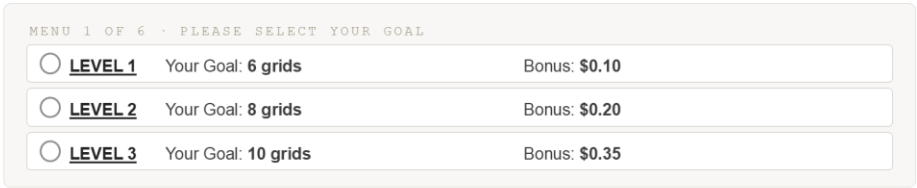
Notes: Omitted: answer-entry training, 45-second practice round, three comprehension questions with corrective feedback. The menu in (b) is the instruction example; experimental menus differ.

Experimental Appendix B.3 — Belief Elicitation and Goal Menus

(a) Attainment beliefs



(b) Baseline goal menu



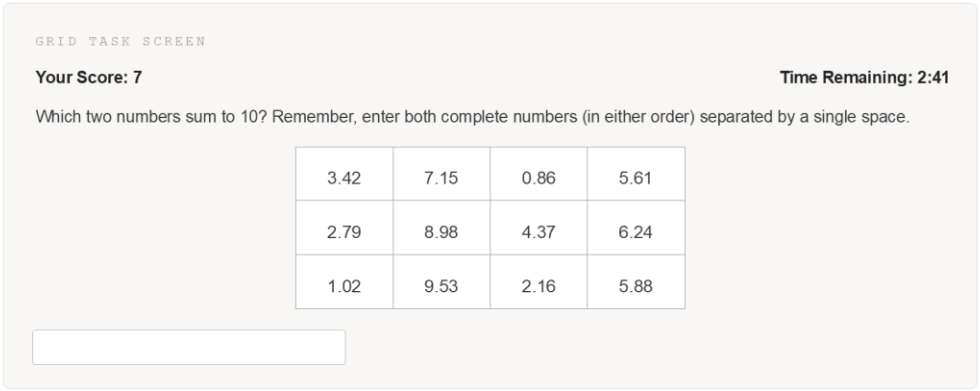
(c) Menu parameters

Menu	Goal thresholds (grids)	Bonuses
Menu 1 (baseline)	6 / 8 / 10	$0.10 / $0.20 / $0.35
Menu 2	4 / 6 / 12	$0.05 / $0.20 / $0.35
Menu 3	8 / 10 / 12	$0.10 / $0.20 / $0.35
Menu 4	10 / 14 / 18	$0.10 / $0.20 / $0.25
Menu 5	6 / 8 / 10 / 16	$0.10 / $0.20 / $0.35 / $0.40
Menu 6	4 / 6 / 8 / 10	$0.05 / $0.10 / $0.20 / $0.35

Notes: Omitted: second slider screen eliciting beliefs for 12, 14, and 18 grids; monotonicity retry. Sliders shown at initial position; menus were presented one at a time in randomized order, one randomly selected for payment.

Experimental Appendix B.4 — Effort Task

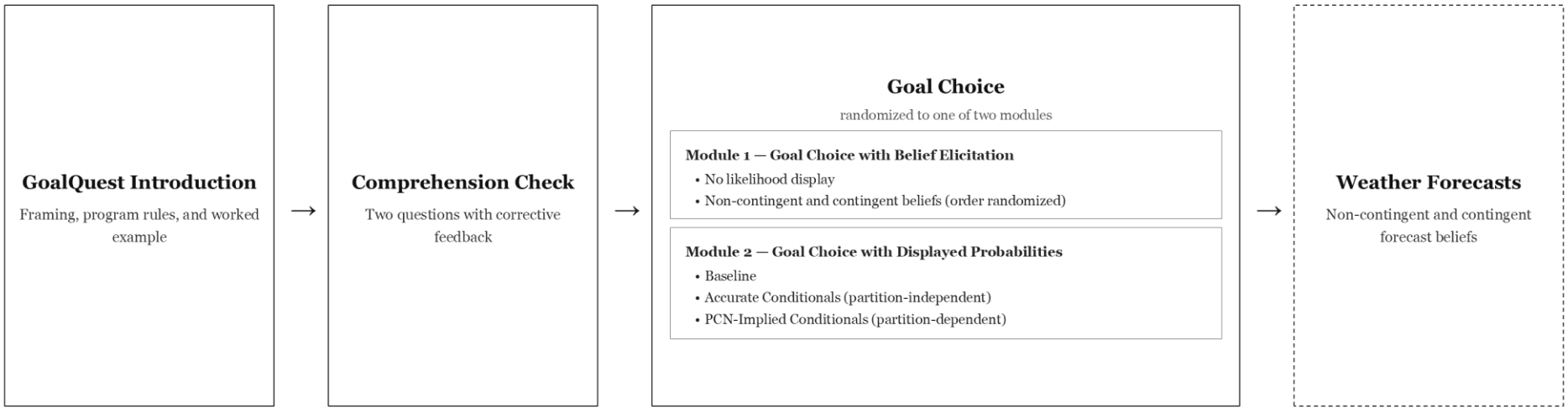
Effort-task interface (4 minutes, repeated grids)



Notes: Omitted: earnings reveal, hypothetical coin-toss bets (loss aversion), competitiveness and financial-risk items. Score, timer, and grid numbers are illustrative ($0.02 base pay per solved grid).

Experimental Appendix C.1 — Schematic Overview

Experiment C: Conditional-belief elicitation and menu-presentation interventions in a hypothetical GoalQuest menu (July 2022 · Amazon Mechanical Turk · analytic N = 815 · hypothetical)



Notes: Omitted: consent, demographics. Both modules used the same menu: goals of 105, 110, and 115 units with rewards of $150, $450, and $900.

Experimental Appendix C.2 — Introduction Screens

(a) Context introduction

SCREEN 1 · CONTEXT INTRODUCTION

Thanks for agreeing to take this survey. For the next series of questions, please imagine that you're a sales employee at ABC, a large US consumer technology company. To encourage greater sales, ABC is set to launch a rewards program called **GoalQuest**. GoalQuest is a proprietary 3-month employee-rewards program used by companies around the world. Next, we will tell you how the program works.

(b) Program rules

SCREEN 2 · PROGRAM RULES

In **GoalQuest**, you are asked to select a sales goal for the next quarter (three months) from a menu of three personalized goal-options.

GoalQuest is an **all-or-nothing** rewards program. So if you achieve your selected goal, you will earn the reward associated with that goal — and only that goal. If you do not achieve your goal, you will earn no reward.

Notes: Omitted: worked example (goals of 56, 64, and 72 units; rewards of $100, $290, and $580), two comprehension questions with corrective feedback.

Experimental Appendix C.3 — Module 1: Goal Choice and Belief Elicitation

(a) Goal choice, no likelihood display

GOAL CHOICE

Now we'd like you to select a GoalQuest goal. To help you make your decision, suppose that over the last several quarters, your sales have been the following (listed in random order):

104, 71, 58, 117, 86, 89, 78, 153, 112, 125, 92, 95, 115, 105 units

Which GoalQuest goal would you choose from the menu below? Please answer as if you were an actual employee participating in this program.

- ○ **GOAL 1**
  Your Goal: Sell **105 units**
  Reward: **$150**
- ○ **GOAL 2**
  Your Goal: Sell **110 units**
  Reward: **$450**
- ○ **GOAL 3**
  Your Goal: Sell **115 units**
  Reward: **$900**

(b) Non-contingent beliefs

NON-CONTINGENT BELIEFS

How likely do you think you are to achieve each goal? Please indicate your percent chance from 0 to 100 percent.

Goal 1 (105 units) 0 — 100

Goal 2 (110 units) 0 — 100

Goal 3 (115 units) 0 — 100

(c) Contingent beliefs

CONTINGENT BELIEFS

Suppose you have a time-traveling friend who travels three months into the future. She returns to the present and tells you that next quarter, you will sell at least 110 units.

Knowing for certain that you will achieve Goal 2 (sell 110 units), how likely do you think it is that you will also achieve Goal 3 (sell 115 units)?

0 — 100

A second screen posed the parallel question for Goal 2 given Goal 1 (sell at least 105 units).

Notes: Omitted: comparison-process report, auxiliary weather-forecasting task. Sliders shown at initial position; elicitation order randomized.

Experimental Appendix C.4 — Module 2: Goal Choice with Displayed Probabilities

(a) Common choice interface

Now we'd like you to select a GoalQuest goal from a real-life menu. To help you make your decision, we will tell you the probability that you achieve each of the goals. Please assume these probabilities are accurate.

Which GoalQuest goal would you choose from the menu below? Please answer as if you were an actual employee participating in this program.

○ **GOAL 1**
Your Goal: Sell **105 units**
Reward: **$150**
‹condition-specific probability statement — see (b)›

○ **GOAL 2**
Your Goal: Sell **110 units**
Reward: **$450**
‹condition-specific probability statement — see (b)›

○ **GOAL 3**
Your Goal: Sell **115 units**
Reward: **$900**
‹condition-specific probability statement — see (b)›

(b) Condition-specific probability displays

BASELINE · NON-CONTINGENT

"You have an **83%** / **74%** / **65%** chance of achieving Goal 1 / Goal 2 / Goal 3."

ACCURATE CONDITIONALS · PARTITION-INDEPENDENT

"You have an **83%** chance of achieving Goal 1."

"If you achieve Goal 1, you have an **89%** chance of also achieving Goal 2."

"If you achieve Goal 2, you have an **88%** chance of also achieving Goal 3."

PCN-IMPLIED CONDITIONALS · PARTITION-DEPENDENT

"You have an **83%** chance of achieving Goal 1."

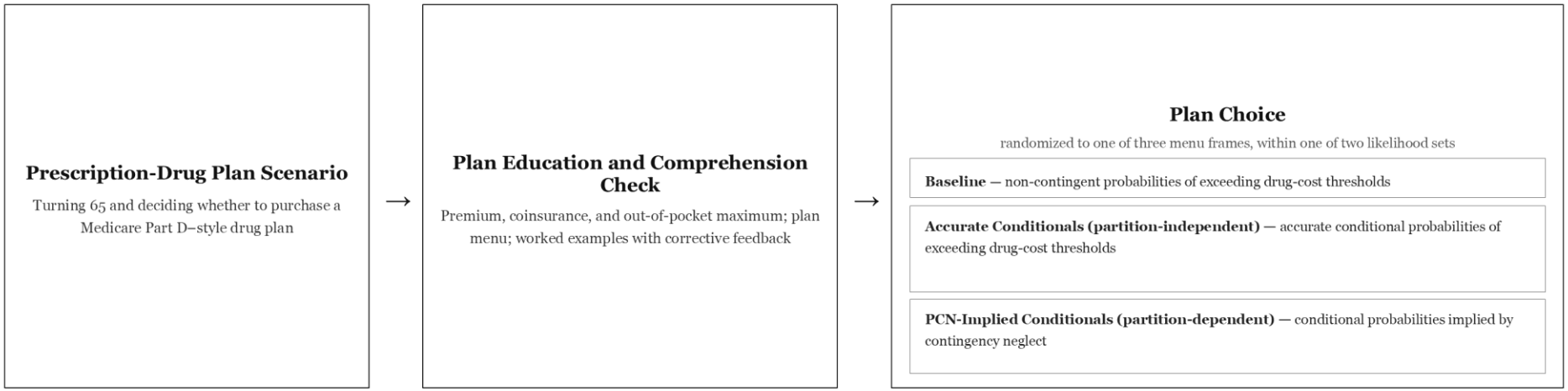
"If you achieve Goal 1, you have a **74%** chance of also achieving Goal 2."

"If you achieve Goal 2, you have a **65%** chance of also achieving Goal 3."

Notes: Each participant saw one menu; goals, rewards, and question wording are identical across arms. Baseline and Accurate Conditionals are informationally equivalent; PCN-Implied Conditionals shows the conditionals implied by full contingency neglect.

Experimental Appendix D.1 — Schematic Overview

Experiment D: Prescription-drug plan choice under three menu frames (August 2023 · Amazon Mechanical Turk · analytic N = 432 · hypothetical)



Notes: Participants received one of two likelihood sets (80/60/40 or 80/63/48), with framing randomized within likelihood set; plan choice was made directly on the framed menu (No Plan · Silver · Gold).

Experimental Appendix D.2 — Scenario and Plan Education

(a) Scenario introduction

SCREEN 1 · SCENARIO

Great. Now please imagine you have just turned 65 and you are finally about to retire. While you already have insurance for your medical care, you are deciding whether to purchase a prescription drug plan from **Medicare Part D**.

This plan would cover any prescription drugs you may need next year (e.g., antibiotics or drugs used to treat high blood pressure, diabetes, high cholesterol, etc.). It does not cover expenses for your partner or dependents. On the next couple pages, you'll see a menu of options varying in annual cost and the amount of coverage.

(b) Plan menu and definitions

SCREEN 2 · PLAN MENU

Here is the menu from which you will be asked to make a selection:

Plan	Annual Premium	Coinsurance	Out-of-Pocket Max
No Plan	$0	—	—
Silver Plan	$640	50%	$7,500
Gold Plan	$1,220	15%	$7,500

The **annual premium** is paid regardless of drug costs; the **coinsurance rate** is the share of drug costs you pay until reaching the **out-of-pocket maximum**, after which the plan covers 100%. Plans have no deductible and cover the same drugs.

Notes: Omitted: three worked examples with corrective feedback. The plan-menu table reproduces the survey's plan parameters (the original menu was displayed as an embedded image).

Experimental Appendix D.3 — Menu Frames and Plan Choice

(a) Common plan-choice interface

Here's some information about next year to help you make a decision.

‹condition-specific likelihood statements — see (b)›

Your drug bill cannot exceed $10,000 even if you select no plan. And within the specified ranges every amount is equally likely for your drug bill (e.g., all values from $1,658 to $10,000 are equally likely).

Which option would you select from the menu below?

- ○ I would select **No Plan**
- ○ I would select **Silver Plan**
- ○ I would select **Gold Plan**

(b) Condition-specific likelihood statements (80/60/40 set shown)

BASELINE · NON-CONTINGENT

"You have an **80%** chance of a drug bill exceeding $0. You have a **60%** chance of a drug bill exceeding $1,280… You have a **40%** chance of a drug bill exceeding $1,657. If this happens, the Gold Plan will be your least expensive option."

ACCURATE CONDITIONALS · PARTITION-INDEPENDENT

"You have an **80%** chance of a drug bill exceeding $0. If your drug bill exceeds $0, you have a **75%** chance of a drug bill exceeding $1,280… If your drug bill exceeds $1,280, you have a **67%** chance of a drug bill exceeding $1,657. If this happens, the Gold Plan will be your least expensive option."

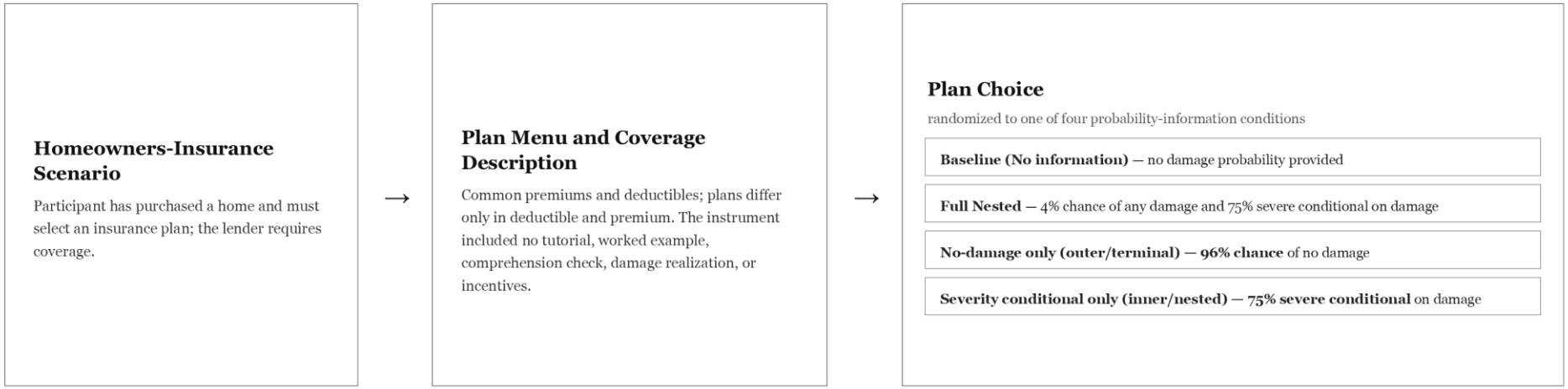
PCN-IMPLIED CONDITIONALS · PARTITION-DEPENDENT

"You have an **80%** chance of a drug bill exceeding $0. If your drug bill exceeds $0, you have a **60%** chance of a drug bill exceeding $1,280… If your drug bill exceeds $1,280, you have a **40%** chance of a drug bill exceeding $1,657. If this happens, the Gold Plan will be your least expensive option."

Notes: Each participant saw one frame; Baseline and Accurate Conditionals are informationally equivalent, while PCN-Implied Conditionals shows the conditionals implied by full contingency neglect. Values shown are one of two randomized likelihood sets (80/60/40; the other was 80/63/48).

Experimental Appendix E.1 — Schematic Overview

Experiment E: Homeowners-insurance plan choice under four probability-information conditions (October 2022 · Amazon Mechanical Turk · analytic N = 302 · hypothetical, not incentive compatible)



Notes: Omitted: consent, gender and age questions, and completion-code screen. The analytic sample comprises 302 responses with nonmissing completion IDs.

Experimental Appendix E.2 — Scenario and Plan Menu

(a) Scenario introduction

SCREEN 1 · SCENARIO

Now we'd like you to imagine that you just purchased a new home and must choose an insurance plan for the next year.

Your lender requires you to have insurance, so **selecting no plan is not an option**.

On the next page you'll see a menu of options from which you can select a plan.

(b) Common coverage description

SHOWN ON EVERY CHOICE SCREEN

Each plan covers the full cost of any damage to your home **after the deductible has been met**. The plans are identical in every way except for differences in deductible and premium.

(c) Common plan menu

COMMON PLAN MENU

Plan	Annual premium	Deductible
Basic Plan	$616	$1,000
Medium Plan	$716	$500
Premium Plan	$803	$250

The coverage description and plan menu were repeated on each condition-specific choice screen, immediately below that condition's probability statement.

Notes: Premiums and deductibles are the survey's own plan parameters. Response-option labels are standardized to "Medium Plan"; one condition screen used "Middle Plan" in its response options.

Experimental Appendix E.3 — Probability Conditions and Plan Choice

(a) Common plan-choice interface

‹condition-specific probability statement — see (b)›

‹common coverage description and plan menu — see E.2 (b) and (c)›

Which home insurance plan do you think you would choose?

- ○ Basic Plan
- ○ Medium Plan
- ○ Premium Plan

The choice page repeated the coverage description and the plan menu from E.2 (c) beneath the condition's probability statement.

(b) Condition-specific probability statements

BASELINE (NO INFORMATION) · NEITHER PROBABILITY

No probability statement appeared before the choice question.

FULL NESTED · OUTER AND CONDITIONAL

"To help you make a decision, suppose your home has a **4 percent chance of at least some damage** over the next year.

If your home does suffer damage, there's a **75 percent chance the damage will be severe (more than $2,500).**"

NO-DAMAGE ONLY · OUTER/TERMINAL

"To help you make a decision, suppose your home has a **96 percent chance of no damage** over the next year."

SEVERITY CONDITIONAL ONLY · INNER/NESTED

"To help you make a decision, suppose that if your home does suffer damage over the next year, there's a **75 percent chance the damage will be severe (more than $2,500).**"

Each statement appeared immediately above the plan menu on the choice page; wording follows the fielded screens.

Notes: Each participant saw exactly one probability condition with an otherwise identical plan menu. The four presentations are *not* informationally equivalent, but they are decision-equivalent: the Basic Plan remains benchmark-optimal over the maintained range of beliefs and risk preferences, so the same plan is optimal in every condition. Full Nested supplied both the outer and the conditional probability; No-damage only the outer probability; Severity conditional only the inner probability.