Partition at Your Own Risk:

Evidence on Risk-Taking Prevalence and Motives from the Field

Saurabh Bhargava
Carnegie Mellon University
sbhar@andrew.cmu.edu

Timothy Hyde
University of Pennsylvania
thyde@upenn.edu

This Version: March 2024

Abstract

Despite its centrality for economic theory, welfare analyses, and policy formulation, attempts to clarify risk-taking motives in the field are often impeded by decision complexity, an inability to observe perceived risk, and limited generalizability. We present evidence on risk-taking prevalence and motives from unique data describing the decisions, productivity, and post-decision beliefs of 20,133 employees across 34 iterations of a $9.4 million goal-rewards program structurally resembling a financial lottery. Our findings reveal substantial risk aversion and choice heterogeneity markedly exceeding predictions of standard expected utility (EU) benchmarks under plausible risk preferences. Conservative choice produced average counterfactual losses equivalent to 85% of rewards, persisted across employee experience and financial stakes ($69 to $4,500), and was notably more frequent for women, contributing to a 21% gender reward deficit. Prominent departures from EU—biased beliefs, non-linear decision weights, and gain-loss utility—failed to meaningfully improve predictive accuracy. After experimentally corroborating choice patterns from explicit menus of economically equivalent lotteries, we advance and experimentally validate a novel heuristic explanation which presumes risk taking emerges from partition-dependent inference in the context of approximate pairwise comparisons. The heuristic explains substantially more choice in the lab and field than other benchmarks. Subsequent experiments demonstrate how heuristic choice could resolve seemingly contradictory empirical insurance puzzles involving excess demand in low-risk settings and inadequate demand in high-risk settings. The findings imply ostensible anomalies in risk-taking prevalence, heterogeneity, and gender disparity across a broad class of decisions may reflect heuristic choice rather than heterogeneous risk preferences and/or perceptions.

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 Reese-Jones, Silvia Saccardo, Emmanuel Saez, Peter Schwardmann, Justin Sydnor, Lowell Taylor, Richard Thaler, Oleg Urminsky and seminar participants at UC Berkeley and Carnegie Mellon 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

Economists have long sought to understand the motives for financial risk taking. Clarifying such motives has profound implications for economic theory, consumer welfare analyses, and the optimal design of programs/policies in domains such as insurance, contract design, financial markets, and consumer protection. From the perspective of Expected Utility Theory (hereafter, EU, or the standard model), the dominant framework in economics for understanding risk-taking, risk aversion among fully informed, utility-maximizing, decision-makers (DMs) reflects the diminishing marginal utility of wealth generated by the concavity of the utility function (von Neumann and Morgenstern, 1947). However, the empirical evidence on decisions under risk (and uncertainty) gives rise to at least three ostensible puzzles with respect to the standard framework.[1] First, economic decisions often imply risk preferences directionally inconsistent with the standard model. For example, in the lab, researchers have documented a degree of risk aversion in small-to-medium sized gambles (e.g., Holt and Laury, 2002) implying implausibly high aversion to risk at larger scales (Rabin, 2000). And in the field, researchers have catalogued instances of excessively high demand, relative to standard model predictions, in low-risk insurance markets (e.g., home insurance) and excessively low demand in high-risk markets (e.g., elderly prescription drug coverage).[2] Second, variation in risk preferences and perceptions have often been seen as insufficient to explain observed heterogeneity in decision outcomes.[3] Last, researchers have relatedly documented systematic differences in risk taking across subgroups, such as gender, not easily explained by differences in preference-based risk aversion (see Niederle, 2017).

In recent decades, researchers have proposed several departures from the standard model to explain risk taking through channels such as biased beliefs, non-linear decision weights (e.g., Kahneman and Tversky, 1979; Prelec, 1998), or gain-loss utility (e.g., Kahneman and Tversky, 1979; Koszegi and Rabin, 2006; Gul, 1991; Loomes and Sugden, 1982). For instance, risk averse choice could stem from systematic overestimation of risk, disproportionate weighting of unlikely outcomes, or an aversion to unanticipated out-of-pocket expenses. It could also emerge from non-standard processes receiving less attention in economics such as heuristics, limited attention, affect, cognitive processes, or hormones (see Kusev et al., 2017; Fox et al., 2015). In practice, research aiming to elucidate risk-taking motives in the field usually investigates settings like insurance, betting markets, or game shows. Such inquiry is,

---

[1] When referring to risk in the paper, we typically mean risk and uncertainty.

[2] Several papers have documented the inconsistency between standard benchmark predictions and observed insurance demand (see Barseghyan et al., 2018). Relative to benchmarks, Abaluck and Gruber (2011) and Heiss et al. (2013) finds insufficient demand for prescription drug coverage; Sydnor (2010) finds excess demand for home insurance; other studies find sub-optimal demand for employer-sponsored health plans (e.g., Handel, 2013, Handel and Kolstad, 2015; Bhargava et al., 2017).

[3] Researchers have asserted that neither heterogeneity in risk (Cohen and Einav, 2007) or risk and risk preferences (e.g., Cutler and Zeckhauser, 2004; Barseghyan et al., 2013) can explain variation in insurance demand. In extensive experimental analyses, Jaspersen, Ragin, Sydnor (2022) find only modest correlation between risk attitudes and insurance demand.

however, often impeded by decision complexity (e.g., consumers may lack full understanding of how to evaluate insurance contracts), an inability to observe subjective risk perceptions, and limited generalizability (e.g., game show contestants may be swayed by their unique environment).

We address these challenges with data from a setting—an employee goal-reward program called GoalQuest© (GQ)—uniquely positioned to clarify our understanding of financial risk taking. The attractiveness of the setting derives from the program's distinct structure and the rare access to contemporaneous perceptions of risk to which we were afforded. Specifically, GQ was designed by a consultancy to improve employee productivity across functions such as sales, customer service, and retention at its (predominantly large North American) client firms through a paradigm informed by behavioral science. At the onset of each one-to-three month GQ program, participating employees privately chose a productivity goal from a standardized menu of three options (g_1, g_2, g_3), personalized, when possible, based on prior performance. Critically, goals were associated with an all-or-nothing reward (e.g., selecting g_3 but only attaining g_2 yielded no reward) denominated in points redeemable for non-monetary prizes at a predetermined rate. To promote ambitious choice, menus typically featured linearly increasing goals (e.g., 100 units, 110 units, 120 units) and non-linearly increasing rewards (e.g., \$100, \$300, \$600), such that g_3 should have maximized expected value (EV) for most well-informed employees. The setting was further distinguished by the consultancy's willingness to temporarily institute an enhanced onboarding module that captured an employee's perceived likelihood of attaining each goal immediately *after* goal selection.

The goal-reward structure and transparency of beliefs permit us to effectively interpret goal choice as a decision between "nested" lotteries—that is, ascending lotteries characterized by a common source of risk such that the winning outcomes of a less risky option subsume those of a riskier one. The setting also affords investigation of typically obscured belief-based motives for risk taking in a context promising high generalizability. Such generalizability derives from demographically and occupationally diverse decision-makers who engage a simple, standardized, menu across varying financial stakes (\$69 to \$4,500) in a program with near-complete participation. The resulting dataset comprised the decisions and beliefs of 20,133 employees across 34 GQ programs and \$9.4 million in rewards—we replicate analyses on an additional decision-only sample of 15,345 employees and \$8.2 million in rewards.

We begin by describing insights from our analyses as to the prevalence of risk taking. A primary finding is to document substantial risk aversion, as nearly one-half of employees selected a goal lower than that predicted by a baseline risk neutral EU benchmark under (imputed) rational expectations. For those whose performance exceeded the low-goal threshold, conservatism led to an average counterfactual loss of \$139, or 85% of the average realized reward. Only 45% of employees chose the benchmark optimal goal, a rate that did not meaningfully vary across reward size, tenure, or salary and was not driven by outlier programs. Employees also exhibited far more heterogeneity in choice (0.35) than predicted by the benchmark (0.75), as indicated by a comparison of Herfindahl-Hirschman Index values. The analyses also revealed substantial gender differences in risk averse choice, as women were 33% more likely than men to select a goal lower than that maximizing expected value. We estimate that this gender gap in choice accounted for roughly one-half of the 21% female deficit in realized rewards.

Our analyses also provide new evidence as to the motives underlying risk taking. This evidence emphasizes the explanatory limits of standard and prominent non-standard explanations from the literature. For example, we find that adopting an EU benchmark with CARA utility with utility-based risk preferences within the interval, $r \in [0.0003, 0.005]$—a range whose upper bound indicates a degree of risk aversion so severe as to imply the rejection of a (∞, -$175) fair gamble—does not improve predictive accuracy relative to the risk neutral benchmark. While assuming severe utility-based risk aversion moderately reduces the share of choice characterized as conservative, it increases the choice share characterized as aggressive by a roughly offsetting degree. A similar challenge arises when attempting to explain conservative choice with CRRA utility across plausible levels of lifetime wealth. Utility-based risk preferences also cannot explain the substantial choice heterogeneity and gender gap in conservative choice we observed—notably, 41% of employee decisions could not be rationalized by *any* value of r within the plausible interval. These patterns parallel the puzzles of conservative and excessively heterogeneous choice, relative to standard benchmark models, routinely found in the economics literature.

Leveraging access to data on employee beliefs, we proceed to investigate three prominent behavioral explanations for risk taking—systematically biased beliefs, non-linear decision weights, and gain-loss utility. We find that they too fail to meaningfully improve predictive accuracy relative to standard benchmarks, nor do they explain observed heterogeneity or the gender gap in conservatism. For example, while systematic underconfidence regarding high goal attainment (or relative overconfidence about low goal attainment) could theoretically explain a propensity towards conservative choice, data on employee beliefs suggests substantial employee *overconfidence* in both relative and absolute expectations of high goal attainment. And contrary to the consensus among economists (see Bandiera et al., 2022), we observe comparable levels of overconfidence across men and women, rejecting differential bias in beliefs as an explanation for gender differences in choice. As another example, while aversion to the prospect of foregoing a potential reward could theoretically deter risky goal choice, across sixty tested gain-loss utility models—spanning credible utility specifications, reference points, and loss aversion parameters—few discernably improved explanatory power relative to a subjective EU baseline. Across all the standard and non-standard benchmark models we considered, including those allowing for heterogenous risk preferences, none explained more than 59% of employee choice and all predicted less choice heterogeneity and a smaller gender gap in conservatism than we observed.

We administered two online studies for additional evidence on risk-taking motives and to address potential confounds from the field. The first study asked participants to select goals from successive goal-reward menus resembling GQ in the context of an incentive-compatible puzzle-completion task. The paradigm permitted us to observe multiple risky decisions per participant in a setting where we could confirm understanding of program rules, explicitly denominate rewards in dollars, and investigate choice from menus of strategically varying size. The experiment yielded a pattern of conservative and heterogeneous choice (and overconfidence) mirroring that observed in the field; it also provided a statistically more emphatic rejection of previously tested benchmarks as well as more flexible versions permitting heterogeneous parameters. We additionally found no evidence to support alternative heuristic explanations from the literature involving contextual sorting via cues such as self-perceived ability (Kamenica, 2008) or taste for competition (Niederle and Vesterlund, 2007). While our research design in the field—particularly the timing of elicited beliefs—was intended to abstract away from alternative motives, such as those involving endogenous effort (a more generalized framework in the Appendix shows how effort motives, at worst, could warrant interpreting our characterization of risk aversion as a slight *underestimate*), a second study corroborates substantial risk aversion and heterogeneous choice from menus that explicitly recast goal choice as a decision between economically equivalent lotteries.

Given the explanatory limits of existing benchmark models, we propose a novel heuristic explanation for risk taking in the present setting and possibly far more broadly. The heuristic—which we refer to as Pairwise Partition Dependence (PPD)—presumes a series of theoretically tractable and psychologically well-founded departures from the expected utility framework. Specifically, PPD posits that employees select GQ goals through successive and approximate pairwise comparisons. Crucially, due to the phenomenon of partition dependence—the sensitivity of inference to possibly arbitrary partitions imposed by a decision context (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Fox and Clemen, 2005)— the heuristic predicts that such pairwise comparisons practically lead most employees to underestimate the relative likelihood of high goal attainment, resulting in greater conservatism and heterogeneity in choice than predicted by unbiased evaluation.

As a concrete example, consider an employee deliberating between g_2 and g_3 (having ruled out g_1). The heuristic stipulates the employee will pairwise compare the goals in the context of three decision-relevant partitions of the state space: attaining neither goal, attaining g_2 but not g_3, or attaining both goals. Due to the low relative likelihood of realizing the middle partition (an expected feature of pairwise comparisons in menus with three or more options), partition dependence predicts systematic overestimation of its perceived likelihood, heightening the subjective attractiveness of g_2 relative to g_3. Beyond privileging conservative goal choice relative to a standard benchmark, by reducing the perceived economic disparity between goals, the heuristic also predicts greater choice heterogeneity—even prior to allowing for decision error, or possible heterogeneity in the severity of inferential bias or in heuristic adoption. While not discussed previously, the heuristic draws on well-established conjectures from the literature involving relative evaluation, partition-dependent inference, and decision noise.[4]

We sought evidence for the proposed heuristic from a new online study and an assessment of predictive accuracy relative to prior benchmarks. An initial arm of the study elicited participant decisions from representative GQ menus in the context of detailed queries of decision process and beliefs. Beyond, once again, replicating previous choice patterns, the study provided evidence supporting the process assumptions of the heuristic. For example, the experiment affirmed widespread participant use of pairwise comparisons to evaluate available goals and, consistent with partition-dependent inference applied to such comparisons, revealed systematic and substantial underestimation of conditional forecasts (e.g., prob($g_3 | g_2$)) relative to the same likelihoods implied from non-contingent elicitations (e.g., prob(g_3)/ prob(g_2)). (We replicated substantial bias in pairwise inference in the distinct context of the weather). The magnitude of bias strongly predicted participant goal choice even after controlling for non-contingent beliefs. A second arm of the study provided experimental evidence consistent with heuristic choice. That is, when randomized to a GQ menu designed to discourage partition dependence, via display of accurate contingent likelihoods of goal attainment, participants were 48% more likely to adhere to EU benchmarks than from an informationally equivalent baseline menu displaying non-contingent likelihoods. Moreover, response to the baseline menu was indistinguishable from response to a menu encouraging partition dependence via the display of contingent likelihoods adjusted for presumed bias. As final diagnostic evidence, we found that the heuristic accurately predicted significantly more choice in the lab and field than any previously tested benchmark; it also explained most of the gender gap in conservative choice.

We conclude by outlining how the proposed heuristic could help understand economic risk taking beyond employee reward programs. Specifically, we see the heuristic as applicable to economic menus that can be conceptualized as offering a choice between nested lotteries. Beyond contingent employee incentive schemes, such menus are routinely found in settings such as portfolio allocation, options trading, and insurance plan choice. We illustrate the heuristic’s specific applicability to consumer insurance by describing a theoretical framework of heuristic choice from a menu of insurance plans varying only in cost and actuarial cost-sharing. Relative to standard benchmarks, the heuristic predicts systematic bias in insurance demand of a direction and magnitude shaped by structural features of the insurance market. For example, the heuristic predicts excessively low demand in insurance markets with

---

[4] Relative evaluation is integral to economic theories of reference-dependent preferences (e.g., Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Koszegi and Rabin, 2007) and comparative evaluation (e.g., Koszegi and Szeidl 2013; Bushong et al., 2021). Decision noise is a common feature of economic models such as those involving bounded rationality or stochastic preferences and its importance in choice and judgment has been emphasized in recent work by Kahneman et al. (2021).

high baseline loss risk and excessively high demand in markets with low baseline loss risk—predictions aligned with recent empirical analyses of Medicare Part D (Abaluck and Gruber, 2011; Heiss et al., 2013) and home insurance markets (Sydnor, 2010). The heuristic also offers a potential explanation for heterogeneous demand that does rely on implausible variation in risk preferences or beliefs.

A final set of experiments investigated whether the proposed heuristic could help resolve seemingly contradictory puzzles in the empirical insurance literature. The experiments asked participants to make hypothetical insurance choices from stylized menus adapted from Medicare Part D and US home insurance markets. Following the earlier paradigm, the menus strategically varied the framing of prospective risk information to either encourage or discourage pairwise partition dependence. At baseline, participant behavior reflected that observed in the empirical literature—inefficiently low demand for prescription drug coverage and inefficiently high demand for home insurance. Consistent with the heuristic, however, menus designed to discourage partition dependence led to a 39% (prescription drugs) and 35% (home insurance) increase in EU-optimal choice relative to baseline despite the informational equivalence of the menus. An additional experiment traced the decay in heuristic bias towards over-insurance across exogenous increases to baseline loss risk, a dynamic also predicted by the heuristic.

Our research relates to multiple, disparate, literatures in economics. First, we contribute to prior work seeking to clarify the prevalence and motives for financial risk taking in the field (see Barseghyan et al., 2018). In addition to documenting substantial risk aversion and choice heterogeneity in a highly generalizable setting, we leverage rare access to data on perceived risk to assess, and reject, prominent risk taking motives from the literature. Second, in proposing an alternative account for risk taking, the paper joins other studies proposing non-standard, and specifically, heuristic, explanations for menu-based decisions in contexts such as asset allocation (e.g., Benartzi and Thaler, 2007) or insurance choice (e.g., Ericson and Starc, 2012; Bhargava et al., 2017; Jaspersen et al., 2022). Because the PPD heuristic implies potentially substantial inferential error, our findings raise the possibility of biased welfare analyses or inaccurate policy inferences in situations where researchers mischaracterize underlying decision processes. For example, the heuristic alludes to the potential challenges of inferring risk preferences from insurance demand; and by documenting violations of descriptive invariance, the experiments identify a strategy for improving consumer welfare via strategically reframed menus not afforded by traditional analyses. Third, our findings contribute to the literature cataloguing gender differences in risk taking in the lab and field (see Niederle, 2017). Our analyses imply that these differences may reflect gender variation in decision strategy rather than differences in risk attitudes or beliefs.

Finally, we see this study as exemplifying how partition dependent inference can systematically and predictably influence a broad class of economic decisions. In this way, the paper complements the program of Ahn and Ergin (2010) who demonstrate how to parsimoniously incorporate partition dependent beliefs into an axiomatic choice framework and the largely experimental literature describing the influence of partition dependence on choice and inference (see Benjamin, 2019). In hypothesizing that partition bias may naturally emerge when decision-makers make pairwise comparisons from ordered menus, our heuristic implies arguably broader applicability of partition dependence than contemplated by prior studies. And in contrast to some other menu-based heuristics, we speculate that one can assess the implications of the proposed heuristic for market dynamics without detailed knowledge of menu design, suggesting its usefulness for various economic analyses of consumer behavior. Relatedly, to our knowledge, the paper is the first to explicitly document the substantial discrepancy in estimates of conditional pairwise likelihoods across contingent and non-contingent elicitations, a finding of potential relevance for research on financial or health literacy, or the methodological design of elicitations targeting beliefs. Lastly, as partition dependent inference is ultimately a descriptive phenomenon, we interpret it as compatible with non-standard decision processes recently discussed in the field such as selective attention and salience (e.g., Bordalo, Gennaioli, and Shleifer, 2012; 2013) and contingency neglect (e.g., Martínez-Marquina, Niederle, and Vespa, 2019; Sunstein and Zeckhauser, 2010).

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. The firm, founded in 1950, specializes in the design and delivery of a suite of proprietary programs that leverage principles from behavioral science (e.g., non-monetary rewards, goal setting, personalization, symbolic recognition, lotteries, contests, communication, and feedback) to improve employee, firm, and consumer engagement. As of 2021, BIW had engaged 6 million individuals across 144 countries through its various products. As of the same date, according to third-party estimates, the firm had approximately 1,500 employees and annual revenues between $500 million to $1 billion. 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.[5] As of 2018, BIW had administered over 1,000 GQ programs to over 1 million participants at firms primarily in the United States, Canada, and Europe since its 2001 inception. While marketed as a sales incentive program, our data indicate that the program has serviced a significant share of employees engaged in customer service and retention (e.g., call centers) across a diversity of sectors (e.g., communication, health care, manufacturing, financials, consumer discretionary).

[5] The World Intellectual Property Organization Publication Number associated with GQ is WO 01/13306 A2 (February 2001).

2.2 GQ Program Overview

Although GQ programs are administered across many contexts, they share a standardized structure that permits comparability across programs. Participation entails three phases: an enrollment period (employees selected a goal), a performance period (employees logged performance towards their goal), and, for those achieving their selected goal, a reward redemption period. During the initial phase, employees were directed to an online portal where they proceed through a simple web-flow.[6] The web-flow itself consists of three parts: a program overview, an explanation of program rules, and goal selection. Employees 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).[7] BIW promotes the program as having a 98 or 99 percent participation rate among eligible employees.[8]

In 2014, we asked BIW to implement an enhanced enrollment process to elicit additional data from employees, namely their beliefs regarding goal attainment. Enhanced enrollment added a fourth phase to the enrollment process: immediately after selecting their goal, respondents were prompted to complete a brief survey. The survey asked employees to estimate their perceived likelihood of goal attainment for 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.

Following goal selection, employees transitioned to a 30- to 90-day performance period during which they attempted to achieve their selected goal. In most programs, participants were able to log onto the website to check their progress or to remind themselves of their selected goal.[9] At the close of the performance period, employees who attained their goal exchanged reward points for a reward in the GQ marketplace. The non-monetary rewards included major electronics (e.g., flat-screen television), event packages, vacations, household items (e.g., luggage), or recreational items (e.g., golf clubs). Employees were educated as to the approximate conversion rate between points and the dollar value of the associated rewards during program marketing; for many programs, employees were familiar with the conversion rate through other BIW programs.

[6] While the design of the enrollment portal was standardized across programs at any point in time, its design evolved over time.
[7] In some programs, the goal selection period may have briefly overlapped the performance window.
[8] While we cannot directly verify participation statistics, high participation rates are plausible due to marketing and communication during the pre-period, the administrative ease of enrollment, and often-valuable rewards.
[9] According to BIW, most programs provided data on intermediate performance to employees. In some programs, intermediate feedback was not technically feasible, or necessary, to track.

2.3 Goal and Reward Structure

We highlight three distinctive structural features of GQ relevant for the present research. First, based on the presumed importance of personal choice and personalization for engagement, GQ required employees to self-select a goal from a personalized menu of options. The personalization reflected the application of some uniform rule to an employee's productivity during an earlier baseline period (e.g., productivity during the prior quarter).[10] Most program menus featured 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, x_b, (e.g., $f(x_b) = 1.05x_b$) and a denotes some increment, potentially itself a function of baseline productivity (e.g., 10 or $0.10x_b$). Employees within a program were usually segregated into a few distinct groups based on comparability in factors such as baseline performance, experience, or job level. While personalization rules could vary across groups—permitting, for example, GQ to assign all new employees to a menu not informed by baseline data—menus within each group were personalized using the same rule. Employees were not given explicit encouragement to select any specific goal via recommendations, defaults, or persuasion.

Second, based on the presumed motivational potency of self-selecting high goals, GQ implicitly encouraged high goal choice by ensuring that high goals were financially attractive. Specifically, in contrast to the additively linear increase in goals, rewards typically increased in non-linear increments. For example, many menus followed the k, $3k$, $6k$ reward structure, where k was approximately 1 percent of an average employee's salary over program duration. Due to the non-linearity of rewards, and their all-or-nothing nature, we estimate that, assuming rational expectations, Goal 3 maximized expected value for 84 percent of employees (Goal 2 maximized EV for 11 percent of employees). Lastly, the rewards associated with each goal were non-monetary, due to a belief that non-monetary rewards would be more motivating than monetary rewards of similar value.

3 THEORETICAL FRAMEWORK OF GOAL CHOICE

We now introduce a theoretical framework to organize our analysis of goal choice. We represent choice as a stylized decision between two simple lotteries varying in their risk and reward. After initially presuming that employees with well-calibrated beliefs select the goal that maximizes expected utility, we consider systematic departures from the standard framework involving biased beliefs, non-standard decision-weights, and gain-loss utility. The framework permits us to characterize observed goal choices as either optimal or conservative relative to predictions of successive benchmark models. In the Appendix

[10] The calculation of baseline performance was jointly determined by BIW and each firm on a program (and often group) specific basis based on considerations of 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.

we present a more generalized treatment of choice in which employees jointly select a goal and commit to some optimal level of costly, but productive, effort. The exercise implies that because of the timing with which we elicited beliefs regarding goal attainment, one can interpret estimates of optimal (conservative) choice under our simplified framework as an upper (lower) bound of estimates one would derive under the more generalized framework.[11]

3.1 Baseline Decision Rule

Our framework describes the decision of a utility-maximizing employee from a simplified menu of two productivity goals associated with all-or-nothing rewards. Due to variance in employee productivity across periods, we represent the two goals by lotteries, $G_n \in [G_h, G_l]$, yielding a reward x_n with some probability s_n, and no reward with some probability $(1 - s_n)$. The high goal has a strictly higher reward, $x_h > x_l$, and lower likelihood of attainment, $s_h < s_l$. Goals are associated with ascending productivity thresholds, $G_h > G_l$, from a common data generating process, such that attainment of G_h implies attainment of G_l. We assume an inter-temporal discount rate of 1, rendering the timing of reward receipt immaterial.

For our baseline benchmark, we assume employees have rational expectations of goal attainment, $\hat{s}_n^r = E(s_n \mid \Phi) + \varepsilon$, where the parameter, Φ, denotes information available to the employee and ε is a normally distributed, mean-zero, error term with constant variance. If $u(.)$ is an always increasing function that maps rewards to utility, then a risk neutral employee selects a goal by solving the following maximization problem:

$$\max_{n \epsilon (h,l)} \mathrm{U}(G_n) = \hat{s}_n^r \, u(x_n)$$

This baseline benchmark implies that an employee will select the low goal if its higher likelihood offsets its lower reward: $\hat{s}_l^r / \hat{s}_h^r > x_h / x_l$.

3.2 Standard Motives for Risk Aversion

We proceed to consider various motives for risk averse choice (that is, a goal lower than that predicted by the baseline benchmark). A first explanation for risk aversion is diminishing marginal utility of wealth in the context of standard expected utility. We incorporate utility-based risk aversion into the benchmark model by adopting a parametric utility function from the constant absolute risk aversion

[11] The intuition for this bounding result (in a two goal setting) is that because we elicit employee beliefs following goal selection, we observe the perceived likelihood of goal attainment conditioned on optimal effort provision given the chosen goal but not counterfactual likelihoods under optimal effort given the non-chosen goal. Consequently, it is possible that ostensibly optimal high goal choices may be conservative and ostensibly conservative low goal choices may be optimal. Such situations could arise if the observed advantage in expected utility of the high goal is offset by an unobserved disadvantage associated with optimal effort under the high relative to the low goal (see Appendix).

(CARA) family. If the parameter, *r*, captures employee attitudes towards risk ($r > 0$ implies risk aversion; $r = 0$ denotes risk neutrality; we ignore the possibility, $r < 0$), we can describe utility by:

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

The choice of a CARA function permits us to represent risk attitudes with a single parameter but implies the irrelevance of prior wealth for an employee's risk preferences. In the Appendix, we consider arguably more realistic utility functions featuring constant relative risk aversion (CRRA) and show that the simplifying CARA assumption does not affect goal characterization.

A utility-maximizing employee with concave utility would now select a low goal if:

$$r > \frac{\ln\left(\frac{\hat{s}_l^r}{\hat{s}_h^r}\right)}{x_l - x_h}$$

The decision rule implies conservative goal choice is not only increasing in the relative expected value of the low goal but also in an employee's utility-based risk aversion. Practically, when characterizing choice with this benchmark we consider risk aversion parameters within a range of plausible values; we further consider benchmark models allowing for heterogeneous risk preferences within this range.

3.3 Non-Standard Motives for Risk Aversion

<u>Non-Standard Beliefs</u> [$\hat{s}_n \neq E(s_n)$]. We next consider the possibility that conservative goal choice reflects non-standard motives such as systematic bias in employee beliefs of goal attainment, $\hat{s}_n$. We denote non-standard beliefs with a goal-specific multiplicative constant, γ_n, such that $\hat{s}_n = \gamma_n s_n + \varepsilon$. Consequently, $\gamma_n > 1$ implies overconfidence while $\gamma_n > 1$ implies under-confidence. A risk averse utility-maximizing employee with subjective beliefs will select the low goal if:

$$r > \frac{\ln\left(\frac{s_l}{s_h}\right) + \ln\left(\frac{\gamma_l}{\gamma_h}\right)}{x_l - x_h}$$

The decision rule implies conservative goal choice increases in relative overconfidence of low versus high goal attainment, γ_l/γ_h.

<u>Non-Standard Decision Weights</u> [$\pi(s) \neq s$]. We proceed to consider whether adopting non-linear decision weights, $\pi(s) \neq s$, into a model of subjective expected utility might help to explain conservative goal choice. Researchers have advanced several probability weighting functions to address apparent violations of the linear weighting assumption of expected utility. We specifically consider the popular

inverse-S shaped function proposed by Prelec (1998): $\pi_n = \exp\left(-(-\ln s_n)^{\alpha}\right)$. In theory, if employees were to underweight high goals more severely than low goals, a non-linear weighting function could help explain conservative goal choice. Assuming non-linear decision weights, the low goal decision rule for a subjective utility-maximizing employee is given by:

$$r > \frac{\ln\left(\frac{\pi_l(\hat{s}_l)}{\pi_h(\hat{s}_h)}\right)}{x_l - x_h}$$

Loss Aversion [$v(x_n, \theta)$]. Finally, we consider the possibility that conservative goal choice may arise from prospective loss aversion in the context of gain-loss utility (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Loss aversion has been advanced as a possible explanation for small- to moderate- scale risk aversion (Rabin, 2000; Rabin and Thaler, 2001) and has been cited as a possible explanation for risk aversion in the field across a range of economic contexts (e.g., Sydnor, 2010). While GQ employees engage prospective, as opposed to actual, losses, expectation-based approaches to gain-loss utility (e.g., Koszegi and Rabin, 2006; Gul, 1991) and research interpreting goals as reference points (Heath, Larrick, and Wu, 1999) suggest the relevance of loss aversion in explaining conservative goal choice. One practical challenge for assessing models of gain-loss preferences is the absence of clear theoretical guidance as to how to specify the reference point, the functional form of gain-loss utility, and the magnitude of loss aversion. With this in mind, we represent gain-loss utility, $v(x_n, \theta_n)$, given some reference point, θ_n, with the following, highly general, value function:

$$v(x_n, \theta_n) = \begin{cases} \eta m(x_n) + u^{+}(x_n - \theta_n), & \text{for x} \geq \theta \\ \eta m(x_n) + \lambda u^{-}(x_n - \theta_n), & \text{for x} < \theta \end{cases}$$

Here, the term $m(x_n)$ is a strictly increasing function capturing consumption utility; u^{+} is a concave utility function capturing gain-loss utility for gains; u^{-} is a convex utility function capturing gain-loss utility for losses; and λ is the loss aversion parameter. The model assumes consumption and gain-loss utilities are additively separable and we specify η as a scaling factor such that $\eta = 0$ collapses to a model with only gain-loss utility.[12] The generalizability of the value function permits us to characterize goal choice across an exhaustive set of potential parameters informed from the literature and the practical configuration of the menu.[13] Given a benchmark model with gain-loss utility, an employee would select the low goal if its

[12] The composite functional representation of gain-loss utility draws on representations from prior work (Sugden 2003; Kobberling and Wakker 2005; Koszegi and Rabin 2006, 2007).

[13] One resource for identifying candidate reference points is provided by recent work that evaluated the success of potential reference points, in the context of gain-loss utility, for explaining menu-based risky choice in the lab (Baillon, Bleichrodt, Spinu 2020). The authors considered prospect-independent (e.g., status quo, the highest reward, the most attainable reward) and prospect-dependent (e.g., the reward of the selected option, the selected option EV) reference points.

higher subjective likelihood offset any deficit in value: $\hat{s}_l/\hat{s}_h > v(x_h, \theta_h)/v(x_l, \theta_l)$. This decision rule implies that conservative choice could arise if high goals were associated with greater prospective losses than low goals.

4 DATA AND SAMPLE CONSTRUCTION

Our analysis of risk-taking relies on program- and employee-level administrative data from BIW. The employee-level data describes goal choice, productivity, beliefs of goal attainment, and demographic detail. The program-level data describes the identity of each firm, the dates of program administration, rules used to assign employees to distinct groups, and details of the GQ menu configuration. In this section, we describe the construction of the primary sample, summarize its key features, and define the variables central to the subsequent analysis.

Primary Sample. Our main evidence on employee behavior and beliefs draws from what we refer to as the *primary sample*. We constructed the primary sample—comprising 20,133 employees across 18 firms, 34 programs, and 232 distinct groups—by applying screening restrictions to an original dataset (n = 38,661) reflecting the universe of data from GQ programs administered between 2014 to 2018 in the US or Canada with enhanced enrollment, at least 100 fully participating employees, and electronically archived data.[14] To arrive at the primary sample, we generated an *expansive sample* (n = 35,478) that excluded roughly 8% of employees from the original dataset for whom a key data field was missing, the data was not consistent, or we inferred incomplete participation.[15] We then restricted the expansive sample to employees providing internally consistent beliefs during enhanced enrollment to create the primary sample.[16] 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.[17] For robustness, we reproduce key analyses for the expansive sample in the Appendix.

---

[14] Data for a small number of programs was not archived by BIW. The size cutoff was necessitated by resource constraints.
[15] 5.2% of the original sample was missing critical data fields, 0.3% of the sample had contradictory award data, and 2.8% of the sample was identified as likely not participating or completing the program based on implausibly low performance reports.
[16] 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.
[17] 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. The comparison suggests that conservatism and sub-optimal choice documented in the primary sample not only exists but may be exaggerated in the expansive sample (we confirm this intuition in Section 5).

Table 1 summarizes overall sample statistics as well as group-level (duration, financial stakes) and employee-level (age, gender, tenure, inferred income) characteristics for the primary sample.[18] On average, we observe data for 592 employees per program (IQR: 208 to 703) and 87 employees per group (IQR: 12 to 103). The groups varied with approximate uniformity across either 30, 60, and 90-day program durations (two outlier programs lasted 45 and 120 days). The distribution of potential reward values was asymmetric, such that 10 percent of employees engaged decisions with rewards averaging $2,150, despite a group-level average of $607 and an employee-level average was $466. Collectively, employees in the primary (expansive) sample had the opportunity to earn $9.4 ($17.5) million in possible rewards. The table also conveys the diversity of the sample across gender, age, and tenure.

Data and Central Measures. Our analysis of choice heterogeneity and efficiency requires data on goal choice, employee beliefs, and indirectly, employee productivity. We describe choice heterogeneity by reporting the distribution of goal choice and by calculating the Herfindahl-Hirschman Index (HHI) implied by the choice distribution. A popular measure of market concentration, when applied to choice settings with n options, the HHI indicates heterogeneity via a standardized index ranging from 1/n (complete) to 1 (none). Table 2 summarizes employee choice, productivity, and goal attainment. Across programs, 44% of employees selected the highest goal with a roughly even split across remaining goals, yielding a 0.35 HHI. The table conveys a correlation between goal choice and two measures of employee productivity, consistent with more productive employees sorting themselves into higher goals (or alternatively, higher goals leading to elevated performance).[19] Figure 1, which presents choice shares across programs and groups, indicates non-trivial variation in choice and an absence of sizable outliers.

We characterize choice efficiency relative to predictions of specified benchmarks via indicators of ex-ante optimality (goals matching benchmark predictions), conservativeness (goals lower than benchmark predictions), or aggressiveness (goals higher than benchmark predictions). This characterization requires data on employee expectations of goal attainment. We draw on two measures of goal attainment beliefs conditioned on goal choice: (1) econometric estimates of rational expectations, and (2) subjective beliefs elicited through enhanced enrollment. To estimate rational expectations, $\hat{s}_{k,i}^{r}$, of employee, *i*, with respect to attaining goal, *k*, we appeal to a strategy routinely used in insurance analyses. The strategy involves constructing employee sub-samples by program group and goal choice and then predicting ex ante attainment likelihoods for each employee and each goal by adjusting the average

[18] Some firms participated in multiple GQ programs sequentially at the same location, so a small number of employees appear in the sample multiple times in different programs.
[19] We did not have baseline data for 16 percent of employees. In most cases, this reflects the lack of past performance data for new employees or programs where performance goals were defined without reference to a baseline.

attainment rate of the sub-sample, excluding the reference employee, by observable covariates.[20] The exercise effectively assumes one can proxy for ex ante rational expectations with ex post average attainment of the appropriate sub-sample. Subjective beliefs of goal attainment, $\hat{s}_{k,i}$, more straightforwardly reflect employee estimates of goal attainment immediately following goal selection (i.e., for tractability, we recoded beliefs of 0% to 1% and from 100% to 99%).

Table 3 summarizes employee beliefs and reveals two notable patterns. First, the table shows a correspondence between goal choice and attainment beliefs, again alluding to an underlying coherence in employee choice. Second, comparing subjective and rational beliefs indicates substantial employee overconfidence regarding future productivity, a point we revisit in the subsequent analyses.

5 CHARACTERIZATION OF CHOICE BY BENCHMARK MODEL

We now characterize employee choice relative to predictions of theoretically informed benchmark models. Specifically, for each benchmark model, we summarize the implied share of optimal, conservative, and aggressive choice; compare observed and predicted choice heterogeneity overall and report the implied gender gap in conservative choice; and report measures of counterfactual loss and moderation of optimal choice across reward size and employee tenure.

5.1 Overview of Risk Taking Prevalence and Heterogeneity

We initially summarize the prevalence and heterogeneity of risk taking under an EV benchmark with rational expectations (Table 3). The characterization exercise implies that 45% of employees chose optimally with a high residual degree of conservative choice, indicating substantial risk aversion. Optimal choice predominantly entailed choosing Goal 3 (87% of employees), followed by Goal 2 (7% of employees) and then Goal 1 (6% of employees). The first panel of Figure 2 shows that most programs and groups feature optimal choice shares between 25 and 75 percent and high residual shares of conservative choice. For employees engaging in conservative choice and attaining at least the low goal, the average reward of $164 implied a counterfactual loss of 46% relative to the $303 average ex ante optimal reward and a loss of 85% relative to the average realized reward (Appendix Figure A2 shows the cumulative distribution of counterfactual loss overall and by goal choice). The propensity towards optimal choice did not vary across reward size quartile or employee tenure.

[20] More specifically, we initially estimated the following leave-out regressions for each employee *i* and goal $k \in [1,2,3]$: $\bar{s}_{k,l,-i} = \alpha + \boldsymbol{Z}\gamma + \pi_l + \varepsilon$. Each regression predicts average group-level attainment for each goal, $\bar{s}_{k,l,-i}$, leaving out employee *i*, as a function of employee characteristics included in vector **Z** (age, tenure, gender) and group fixed effects, π_l. (We estimated regressions at the program level to increase the precision of covariate estimates). We then calculated an employee's rational expectation of attaining goal *k*, as $\hat{s}^r_{k,i} = \hat{\alpha} + \boldsymbol{Z}\hat{\gamma} + \hat{\pi}$.

Employee choices also implied substantially more heterogeneity (0.35 HHI) than predicted by the baseline benchmark (0.75 HHI). The preponderance of data situated below the diagonal in the second panel of Figure 2 indicates most programs and groups exhibited (often significant) excess heterogeneity. With respect to sub-group heterogeneity, women were more apt to select lower goals (i.e., Goals 1 or 2) than men (0.63 versus 0.50), resulting in substantially greater risk aversion under the EV-benchmark (0.57 versus 0.43, yielding a gender gap of 0.14). We can approximate the economic consequences of differential risk taking by estimating the share of the overall gender gap in rewards—women earned 21% less rewards than men—attributable to gender differences in conservative choice. The estimates indicate that had women adopted the same risk profile as men, all else equal, women would have closed the gender gap in rewards by 48%.[21]

5.2 Standard Motives for Risk Taking

We now consider the possibility that conservative and excessively heterogeneous choice relative to the baseline benchmark reflects utility-based risk aversion. We model risk aversion by assuming a CARA utility function with risk preferences within the interval, $r \in [0.0003, 0.005]$. To appreciate why this interval subsumes the range of plausible risk preferences, one can translate the preferences implied by the interval to gambles involving potential losses comparable in magnitude to GQ rewards. For example, consider a simple 50/50 lottery with a chance of losing \$175 (roughly the 25th reward percentile) and a chance of some unspecified gain. The lower bound of the interval, r = 0.0003, implies acceptance of the gamble for potential gains exceeding \$184; it implies acceptance of any 50/50 gamble with a potential loss of \$350 (roughly the median GQ reward) so long as the potential gain exceeds \$391. In contrast to these seemingly plausible risk attitudes, the interval's upper bound, r = 0.005, implies rejection of *any* 50/50 gamble involving a potential loss of \$175 or \$350, even with an infinite potential gain.

As Table 3 conveys, the assumption of modest risk aversion (r = 0.0003) does little to shift the baseline characterization of choice or the implied degree of excess heterogeneity. The assumption of (arguably implausibly) severe risk aversion, r = 0.005, moderately shifts the characterization of choice from conservative to aggressive and implies less excess choice heterogeneity than the baseline EV-benchmark. Severe risk aversion does not, however, meaningfully affect the implied share of optimal choice, the magnitude of counterfactual loss, nor the absence of moderation in choice efficiency by reward size or employee tenure. (Figure 3 depicts the relative insensitivity of characterized choice to reward magnitude with higher granularity). Even allowing for fully flexible risk preferences—i.e., tagging

[21] We estimate the share of the reward gender gap attributable to differential risk taking by regressing realized reward dollars on indicators for gender, conservative choice, and optimal choice. We use the estimated coefficients to predict the additional rewards women could accrue from equalizing their rate of conservative and optimal choice.

every choice that can be rationalized by any risk parameter within the interval, $r \in [0, 0.005]$, as optimal) explains only 56% of goal choice and does not further reduce the implied gender gap in conservatism.[22] The explanatory limits of standard motives are not constrained to CARA utility. Appendix Table A1 shows that assuming constant relative risk aversion (CRRA) utility across a range of potential wealth produces a nearly identical choice characterization to CARA benchmarks.

Figure 4 provides intuition as to how variation in utility-based risk preferences affect choice characterization. The figure, which depicts predicted choice shares for each goal under EU benchmarks across a "super-plausible" interval of risk aversion, $r \in [0.00, 0.10]$, illustrates that significant shifts in characterization largely involve risk preferences outside the previously described interval of plausible risk preferences (shaded). Within this interval, greater risk aversion modestly increases the optimality of low goal choice but decreases the optimality of high goal choice by a commensurate degree—this dynamic helps to explain why the assumption of plausible risk aversion leaves the implied share of optimal choice largely unchanged despite such benchmarks predicting greater choice heterogeneity.

5.3 Non-Standard Motives for Risk Taking

We proceed to consider whether one can explain conservative and heterogeneous choice through prominent departures from the standard EU framework. We begin by assessing the predictive accuracy of a subjective expected utility (SEU) model. While standard economic theory stipulates that variation in risk perception should produce variation in risk taking, in the present context, an SEU benchmark could account for the observed behavior only if employees exhibited systematic bias in their relative beliefs regarding the attainment of high versus low goals. Beyond considering such systematic bias, we consider the explanatory efficacy of models allowing for non-linear decision weights and gain-loss utility.

Biased Beliefs. To assess the possibility that conservative choice reflects systematic bias in perceptions of goal attainment, we characterize choice under an SEU benchmark with plausible risk preferences. As Table 3 indicates, the transition from rational expectations to subjective beliefs does not meaningfully affect the implied share of conservative choice despite modestly improving the share of choice characterized as optimal. Additionally, the adoption of subjective beliefs does not diminish the implied degree of excess heterogeneity relative to analogous benchmarks with rational expectations; only modestly reduces the implied gender gap in conservative choice; and does not reveal notable moderation by reward size or employee tenure (see Figure A3 for a distributional comparison of counterfactual losses by information regime). Even allowance for highly flexible risk preferences, $r \in [0, 0.005]$, implies a sub-

[22] A fully flexible model produces no evidence of significant moderation in efficient choice by employee tenure but does yield some evidence for moderation by reward size. This latter effect is partially mechanical, however, since it is more likely that some r will rationalize two goal choices with larger rewards relative to smaller rewards.

optimal choice share exceeding 40 percent. As with models of heterogeneous risk under rational expectations, incorporating heterogeneous risk in an SEU benchmark cannot explain the implied gender gap in conservative choice.

Table 4 offers insight as to the limited explanatory benefits of incorporating subjective beliefs into benchmark models. The table documents substantial employee overconfidence on average with respect to attaining each of the three goals. More relevantly, the table indicates substantial relative overconfidence with respect to attaining Goal 3 compared to lower goals. While the precise effect of biased beliefs on characterized choice depends on the distribution of relative beliefs of attainment (see Figure A4), the table conveys substantial average bias in a direction consistent with more aggressive, rather than more conservative, goal choice. Not reported in the table, both men and women exhibit similar degrees of absolute and relative overconfidence regarding high goal attainment, helping to explain why gender differences in conservative choice cannot be attributed to gender differences in overconfidence.[23] Comparing the two panels of Figure 4 offers additional perspective as to why transitioning from rational expectations to subjective beliefs only modestly shifts the characterization of choice.

Non-Linear Decision Weights. An additional behavioral departure we consider is the possibility that conservative choice emerges from the influence of non-linear decision weights. To test this possibility, we replace the linear decision weights with the weighting function of Prelec (1998; $\alpha = \beta = 0.65$). Table 5, which compares the choice characterization for non-standard benchmarks to an SEU baseline ($r = 0.0003$), indicates that the assumption of non-linear decision weights does not meaningfully shift the benchmark characterization of choice; nor does it help to explain observed choice heterogeneity or the gender gap in conservative choice. The lack of explanatory improvement is perhaps unsurprising given that for most employees, an inverse-S-shaped weighting function underweights all goal options to a roughly similar degree relative to linear weighting.

Gain-Loss Utility. Finally, we consider the possibility that conservative and heterogeneous choice may reflect prospective loss aversion in the context of gain-loss utility. Given the aspiration to consider all reasonable potential representations of gain-loss utility, we assessed the predictive accuracy of a range of benchmark models reflecting several candidate reference points, θ, functional scaling factors, η, and loss aversion parameters, λ, in the context of subjective beliefs and linear decision weights. Specifically, we considered five prospect-independent reference points: status quo (i.e., $0), the reward associated with the highest probability goal (Goal 1), the highest available reward (Goal 3), the reward associated with the highest goal an employee felt certain to achieve (or $0 if an employee was not certain about attaining any

[23] For example, if one defined overconfidence as the average difference in perceived and actual attainment, men and women were identically overconfident with respect to Goal 3 (both 0.32, $p = 0.73$).

goal), and, for completeness, Goal 2. We additionally considered prospect-dependent reference points including the chosen goal, the expected value of the chosen goal, and in recognition of models of counterfactual regret, the nearest-goal either below or above the chosen goal. We assessed these reference points in the context of composite utility specifications assuming a Kahneman-Tversky power function ($\alpha = 0.88$) for both consumption and gain-loss utility and, given the lack of empirical consensus in the literature, a range of scaling factors, $\eta \in [0, 5]$. And in deference to the breadth of loss aversion parameters contemplated by the literature, we consider $\lambda = 1.5$, 2.25, and 3.0.

Across tested benchmarks, the most predictive gain-loss model—entailing a reference point set at the chosen goal reward, $\eta = 1$, and $\lambda = 2.25$— explained 59% of employee choices (see Appendix Table A4 for prediction rates for all tested models). As reported in Table 5, beyond moderately improving the share of explained goal choice relative to baseline, the most predictively accurate gain-loss model predicts far greater choice heterogeneity and implies a smaller gender gap in conservative choice than previously tested benchmarks (it does not, however, suggest meaningful moderation in choice efficiency by reward size or employee tenure). We additionally considered the possibility that observed behavior may reflect gain-loss utility with heterogenous severity of loss aversion (including the absence of loss aversion). To examine this prospect, we reconsidered the gain-loss utility benchmark after flexibly assigning an employee any $\lambda = (1.0, 1.5, 2.0, 2.5, 3.0)$ that could rationalize their goal choice. The exercise yielded an optimal choice share of 0.71, though we note the imprecision of this estimate, as flexible loss aversion mechanically accommodates two optimal goals for a significant fraction of employees. We revisit the predictive accuracy of the gain-loss benchmark with higher powered tests in the subsequent section.

5.4 Synthesis, Potential Confounds, and Robustness Analyses

Collectively, excepting the gain-loss benchmark with heterogenous loss aversion, attempts to explain employee goal choice with standard EU benchmark models, or common behavioral departures from such models, even allowing for heterogeneous risk preferences, consistently fail to explain over 40% of employee choice, predict less heterogeneity in choice than we observe, and do not explain systematic gender differences in conservative choice. Employee conservatism results in an estimated loss of rewards equivalent to 70 to 100% of the realized reward, or 40 to 50% of the counterfactual reward under ex ante optimal choice. There is no evidence of moderation in efficient choice by reward size or employee tenure. It is possible that a modest allowance for imprecision in decisions or beliefs, particularly imprecision that systematically favors low goal choice, might substantially improve the predictive accuracy of tested benchmarks. However, we find that incorporating non-trivial allowances for noise that favor conservative choice does not result in significant improvements to baseline benchmarks.[24]

Figure 5 provides additional detail as to choice dynamics, under the SEU benchmark (r = 0.0003), across a matrix of subjective perceptions of the high (x-axis) and highest-valued low (y-axis) goal values. The figure suggests an underlying regularity in employee decisions with greater optimality in quadrants where the value of one goal option significantly exceeds the other and lower optimality along the equal-valued diagonal. Despite this regularity, however, the figure alludes to widespread sub-optimal choice. At least 20% of employees choose sub-optimally in nearly every cell with most cells near the diagonal featuring optimal choice shares under 50%. The figure also indicates substantial sub-optimal choice—effectively a strong preference for the low goal—in scenarios where the high goal represents modestly higher expected value. Overall, the figure is consistent with a moderate to strong preference for low goals excepting situations where the high goal offers an overwhelmingly favorable valuation.

Effort Costs as a Confound. Conceivably, conservative goal choice could reflect a motive—such as the convex costs of employee effort—not involving employee attitudes towards risk. While the timing and wording of belief elicitations were intended to abstract away from such a confound (see Appendix), it is conceivable some employees mistakenly interpreted the belief elicitation as inquiring about the likelihood of goal attainment conditioned on optimal effort under that specific goal. For example, an employee might select a low goal to avoid the costly effort required to attain a high goal and when asked to report the post-decision likelihood of attaining non-selected goals, the employee might have instead reported their ex ante beliefs of attainment conditioned on a different expectation of exerted effort. While we interpret such a scenario as unlikely, it would introduce an alternative motive for low goal choice.

While we explicitly address potential confounds experimentally in the subsequent section, we briefly address effort-cost motives and belief confusion through a simple calibration. The exercise involves characterizing optimal choice under a risk neutral SEU benchmark modified to account for effort costs manifest exclusively through percent reductions in hourly wage (notably, effort costs must be convex across ascending goals to explain a systematic preference for lower goals). Adopting a non-parametric approach, we specify a parameter, σ, to represent the baseline incremental effort cost required to achieve Goal 2, relative to Goal 1, as a percent of hourly wage. We represent effort cost convexity with a scaler, $k \geq 1$, such that $k\sigma$ denotes the incremental hourly cost of effort to achieve Goal 3, relative to Goal 2. Appendix Table A2 reports the resulting optimal choice shares across an extreme range of

[24] Allowing for non-trivial decision error favoring low goals only slightly improves the accuracy of standard benchmarks. We implement an allowance for noise by evaluating whether the subjective EU model with moderate risk aversion can rationalize choice for any set of subjective beliefs within a +/- 10 percent range of the self-reported figure. The allowance of a 20 percent error in subjective beliefs only moderately increases the share of choice deemed to be optimal, from 0.50 to 0.54.

parameter values. The table conveys that any plausible representation of convex effort costs reduces explanatory power relative to baseline. Intuitively, given convex and cumulative effort costs, one can easily rationalize the systematic choice of either Goals 1 or 3. To rationalize non-trivial Goal 2 choice, however, one must assume a baseline increment and convexity scaler within a highly—and perhaps implausibly—narrow interval that itself must vary across employees based on variation in subjective beliefs. Even then, any model would predict sharply lower goal choices for employees in longer, relative to shorter, programs—a correlation for which there is no evidence.[25]

Robustness Analysis –Expansive Sample. To assess robustness and generalizability of the findings, we replicate the preceding analysis for standard risk motives using the expansive sample (i.e., the sample inclusive of employees for whom beliefs were not observed). As summarized in Appendix Table A3, choice characterizations in the expansive sample under standard benchmark models assuming rational expectations resemble those of the primary sample but for implying a modestly higher share of conservative choice and a smaller share of optimal choice (the expansive sample, like the primary sample, exhibits far more choice heterogeneity than predicted).[26] The analysis suggests that to the extent the primary sample is not fully representative of the employee population, it modestly *underestimates* the share of sub-optimal and conservative choice.

6 ANALYSES OF POTENTIAL MECHANISMS

We continue to investigate the motives for conservative goal choice through a series of online studies. In the first study, we re-examine prior benchmarks with greater statistical power and assess alternative explanations from the literature in an incentive-compatible context. In a second study, we replicate choice patterns from the field with menus featuring explicit lotteries. And in a third study, we evaluate a novel, heuristic, explanation for GQ goal choice after which we assess its predictive accuracy in the lab and field.

6.1 Additional Tests of Mechanisms via Online Goal-Reward Paradigm (Experiment A)

Overview. We administered Experiment A in May 2019 on the Qualtrics platform to 407 employed US adults recruited from Amazon Mechanical Turk. The online study asked participants to complete a brief puzzle completion task in the context of an incentive-compatible goal-reward paradigm lasting a few minutes. The paradigm resembled GQ but for comprehension checks, dollar-denominated

[25] A regression of an indicator for non-high goal choice on a linear index of program length in days with standard errors clustered at the program level yields a non-significant coefficient estimate, b = 0.001 (p = 0.20). The analogous regression for conservative goal choice (under a subjective EV benchmark) yields a near identical coefficient estimate, b = 0.001 (p = 0.19).
[26] To characterize choice under rational expectations in the expansive sample, we adhere to the previously adopted strategy but for excluding unobserved demographic variables in the regression estimates of beliefs.

rewards, and multiple decisions per subject. The study also captured decision-relevant information including performance likelihood forecasts, loss aversion imputed from hypothetical gambles, and self-reported assessments of relative ability and taste for competition.[27]

We implemented the goal-reward paradigm by first informing participants that they were about to engage in a timed effort task where they could earn financial rewards for solving a series of grids. Solving a grid entailed finding the unique pair of numbers whose sum equaled 10 across a 3 x 3 matrix of single digit numbers. After an opportunity for practice, we formally introduced the goal-reward paradigm (labeled "GoalQuest"), explaining that participants would have four minutes to solve as many grids as possible and that they could earn an all-or-nothing reward by attaining their self-selected performance goal. After a comprehension test, we asked participants to select a goal from six distinct menus presented in succession—a design intended to increase statistical power and facilitate tests of mechanisms (we explained that a single menu would be randomly selected to calculate their reward). The six menus included a baseline that resembled GQ in attainment difficulty (6 grids, 8 grids, 10 grids) and non-linearly increasing rewards ($0.10, $0.20, $0.35); three additional menus modified the baseline by either varying overall difficulty or the relative generosity of the high reward; two menus expanded the baseline menu to four choices by adding a relatively unattractive high or low goal option. Participants then completed the task so we could determine their payment. Excluding data for incomplete and/or inconsistent beliefs, 277 remaining participants made 1,662 goal choices.

<u>Baseline Comparison of Lab and Field</u>. The experiment produced baseline patterns of judgment and decision-making resembling the field. Specifically, average baseline goal choice (0.34, 0.28, 0.38) and beliefs of attainment likelihood (0.80, 0.66, 0.51) resembled averages from the field (choice: 0.29, 0.27, 0.44; beliefs: 0.78, 0.69, 0.63). Based on realized performance, participant beliefs also implied overconfidence, though less severely so than the field. Most notably, characterized baseline choice from the lab (optimal: 0.50, conservative: 0.45) was similar to the field (0.50, 0.48) under a risk neutral SEU benchmark. We interpret the overall correspondence in choice, beliefs, and characterization across lab and field as inconsistent with program confusion, managerial signaling, or reputational concerns exerting outsized influence in the field, as such motives were eliminated or diminished in the lab. Comparability across lab and field also supports the usefulness of the lab paradigm for assessing mechanisms.

<u>Reassessment of Prior Benchmark Models</u>. Table 6 characterizes the efficiency of participant choice under previously considered benchmark models. Beyond reporting the share of participants whose full set of six choices aligned with predictions of each benchmark, the table also reports optimal choice

[27] We elicited beliefs for select performance thresholds and used these to interpolate/extrapolate a more detailed distribution. We measured self-assessed relative ability to complete puzzles and relative taste for competition on five-point scales.

shares allowing for some decision error in the form of only five of six adherent choices. No previously tested benchmark (excluding those with heterogenous parameters) explained more than 18% of participant choices, a rate rising to 31% with decision error. The experiment, however, also provided additional insight, with far greater statistical power, into the accuracy of gain-loss utility benchmarks with heterogenous loss aversion. Specifically, incorporating personalized estimates of loss aversion into the most successful gain-loss benchmark yielded an optimal choice rate of 19% while a fully flexible version of the same benchmark—one assigning each participant the most predictively accurate, $\lambda \epsilon$ [1.0, 1.5, 2.0, 2.5, 3.0]—yielded an optimal choice rate of 29%.[28] Overall, the experiment provides little evidence for previously considered, standard or non-standard, risk-taking motives.

Contextual Sorting Heuristics. The experiment afforded an opportunity to assess the predictive accuracy of two heuristic choice strategies, involving contextual sorting, from the literature that could not be examined in the field. The strategies presume employees, otherwise unsure of what goal to select, heuristically chose a goal whose relative position in the ordered-menu corresponded to their perceived standing in some choice-relevant distribution such as ability or taste for competition (e.g., Kamenica, 2008; Niederle and Vesterlund, 2007). As indicated in the table, we find no support for either heuristic.[29]

6.2 Replication of Choice Patterns with Explicit Financial Lotteries (Experiment B)

As a final test to rule out potential confounds in the field, we sought to corroborate choice patterns in the explicit context of nested financial lotteries economically equivalent to GQ goals (i.e., lotteries of ascending risk and reward characterized by a common source of risk such that the set of winning outcomes of a less risky option subsume those of a riskier option). Accordingly, we administered Experiment B in December 2023 to 243 US adult participants from Amazon Mechanical Turk. Across two conditions, participants were asked to make decisions from a hypothetical employee rewards program. The first condition introduced participants to the real-life GQ paradigm, tested comprehension of the paradigm, and presented them with a stylized menu representative of (percent-denominated) GQ programs from the field (i.e., goals: 105 units, 110 units, 115 units; rewards: \$150, \$450, \$900).[30] In contrast to the field, however, the menu explicitly conveyed the likelihood of attaining each goal using figures approximating empirically-informed averages (83%, 74%, 65%). As such, the menu strongly

---

[28] For roughly 20% of participants, we could not calculate a personalized loss aversion parameter. We ignored such participants from the optimal share rate calculation for that benchmark.

[29] We tested the ability-sorting heuristic by mapping relative assessments of ability to predicted goal choice by menu position (e.g., high relative ability predicts high goal choice) and then comparing actual and predict choice. We used a similar procedure to test the taste-for-competition heuristic.

[30] The menu was representative of percent-denominated GQ programs (i.e., those with rewards expressed as a percent of baseline). To generate rewards we applied the modal rewards ratio (1-3-6) to the median Goal 1 reward (\$150, after rounding). Goals reflect a 5-10-15 percent increase relative to a baseline of 100, reflecting the mean/median/ modal configuration of percent-denominated program. Average rewards in such programs were higher than the global program average.

implied the optimality of Goal 3 under an EV or EU (with plausible risk preferences) benchmark. The second condition tested for risk taking in a context even more explicitly resembling a lottery paradigm. Participants in this condition were informed their employer had adopted "RewardQuest" (hereafter, RQ), a program designed to reward employees for prior departmental performance by permitting them to select from a menu of three reward lotteries. While the lotteries were identical in risk and reward to those in the first condition, to emphasize their nested and random structure, each lottery was associated with a points threshold and participants were told that lottery success would depend on the points accrued from the random spin of an electronic wheel (as in first condition, the likelihood of success was also displayed explicitly). After a comprehension test, participants selected their preferred reward lottery.

Overall, participants from Experiment B exhibited the same patterns of heterogenous and risk averse choice as in the field. Specifically, the experiment revealed a diverse pattern of choice across both GQ (0.26, 0.44, 0.30) and RQ (0.30, 0.45, 0.25) menus. These choices implied substantial conservatism relative to optimal choice under EV/EU benchmarks, with the RQ menu producing modestly (but non-significantly) greater EV-conservatism (0.75) than the GQ menu (0.70; p = 0.41).

6.3 Heuristic Explanation for Risk Taking - Pairwise Partition Dependence (PPD)

What might explain heterogeneous and risk averse choice in the lab and the field? Given the challenges of explaining observed choice through existing benchmark models, we propose a novel heuristic explanation for risk taking informed by our reading of the literature and exploratory pilot studies investigating the phenomenology of participant choice. The proposed explanation, which we refer to as the Pairwise Partition Dependence (PPD) heuristic, broadly stipulates that a DM selects an option from a menu of simple, nested, lotteries through a succession of approximate (adjacent) pairwise comparisons. Critically, the heuristic presumes that such pairwise comparisons can lead to a potentially substantial inferential bias due to the phenomenon of partition dependence. When applied to the GQ menu, the heuristic predicts that employees underestimate the relative likelihood of high goal attainment and select more conservatively—and heterogeneously—than standard benchmarks would predict.

Model Setup. We outline the PPD heuristic more formally in the context of GQ by returning to our earlier framework where we represent goal choice as a decision from a menu of two simple, nested, lotteries ordered from low to high risk (G_l, G_h), such that G_n yields reward, x_n, with some probability s_n and 0 otherwise. The high risk goal has a strictly higher reward, $x_h > x_l$, and lower likelihood of attainment, $s_h < s_l$. Goals are ascending, $G_h > G_l$, so that attainment of G_h implies attainment of G_l. The heuristic specifies that the DM compares the goal pair by evaluating whether the expected marginal gain from the high relative to the low goal outweighs the potential loss of the low-goal reward. In service of this pairwise comparison, the DM partitions potential outcomes into three decision-relevant states demarcated by the two goals: a first state, S_L, specifies outputs below G_l for which goal choice is immaterial; a second state, S_H, specifies outputs for which the high goal reward exceeds that the low goal reward; a third state, S_M, specifies outputs between the two goals and for which the low goal reward exceeds the zero payoff of the high goal. We denote a DM's perceived likelihood of state, S_K, by $\widehat{\varphi}_K$.

Pairwise Decision Rule. The heuristic specifies that a risk-neutral DM will select the high goal if the following decision rule is satisfied:

$$\widehat{\varphi}_{L+}\left[\widehat{\varphi}_{H|L+}\Delta x_{h,l} - (1-\widehat{\varphi}_{H|L+})x_l\right] > \omega$$

The parameter, $\widehat{\varphi}_{H|L+}$ denotes subjective belief of high goal attainment conditioned on at least low goal attainment, $(1-\widehat{\varphi}_{H|L+})$ denotes subjective belief of the complement, $\Delta x_{h,l}$ is the difference between the high and low goal reward, x_l is the low goal reward, and ω denotes some decision error. While the formulation emphasizes the relative nature of the evaluation, we can simplify the rule to: $\widehat{\varphi}_H\Delta x_{h,l} - \widehat{\varphi}_M x_l > \omega$. Absent inferential bias, the rule simply restates the risk neutral utility-maximizing proposition but for the allowance of decision noise.

The heuristic, however, presumes that pairwise comparisons are subject to partition dependence, a phenomenon describing the sensitivity of judgment to the potentially arbitrary partitions imposed by the decision context (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Fox and Clemen, 2005). Specifically, the heuristic contends that individuals form partition-dependent beliefs in the context of the three partitions introduced by the pairwise comparison (S_L, S_M, S_H). Without loss of generality, we represent partition dependent beliefs as a convex combination between partition-independent beliefs and a naïve prior that assigns an equal probability to each partition, following a strategy employed by Fox and Clemen (2005). Accordingly, we define a parameter, $\theta \in [0, 1]$, that specifies the degree of partition bias ranging from an absence of bias, $\theta = 0$, to full bias, $\theta = 1$, so that: $\widehat{\varphi}_K = \theta/\pi + (1-\theta)\varphi_K$. Here, π indicates the number of decision-relevant partitions, which under pairwise evaluation—or in the case of a menu with only two options—is fixed at three. Under full partition bias and renormalization of the noise parameter, the heuristic implies that the DM will select the high goal if $(\Delta x_{h,l} - x_l) > \omega$.

While we have restricted discussion thus far to a menu of two options, one can straightforwardly adapt the heuristic to larger menus, such as the ones featured in GQ by specifying a starting point and stopping rule. For tractability, we assume that DMs heuristically engage menus with more than two risk-ordered options by successively comparing proximal pairs of options beginning with the low-risk option, stopping any time a riskier option is rejected. As such, we assume GQ employees initially compare Goals 1 and 2, and either accept Goal 1 or proceed to compare Goals 2 and 3.

To illustrate how the proposed heuristic might help to explain conservative and heuristic goal choice in GQ, consider an example in which an employee must decide between low goal, G_l = ($300;

0.80), and high goal, G_h = ($500; 0.60), a scenario credibly resembling the decision employees in the field faced between Goals 2 and 3. A risk neutral EU-maximizing employee would decide in favor of the high goal since the expected marginal gain from the high goal [$120 = ($\Delta x_{h,l}$ = $200) x ($\widehat{\varphi}_H$ = 0.60)] comfortably exceeds the expected marginal loss associated with the low goal [$75 = $300 x ($\widehat{\varphi}_M$ = 0.20)]. In contrast, the proposed heuristic would predict low goal choice for any θ exceeding 0.64, the parameter value marking the indifference point between the goals. One would expect in menus with three or more options, pairwise comparisons should—by construction—produce narrow middle partitions, S_M, leading to partition-induced overestimation of perceived likelihood and a bias towards conservative choice.

Figure 6 provides graphical intuition for the application of partition dependence to pairwise inference by depicting the distortion in perceived goal attainment under heuristic and standard choice. The shaded region between the two CDFs depicts inferential bias under heuristic evaluation. It is worth noting that the heuristic does not inherently bias DMs toward less risk. Instead, the direction of bias depends on the economic structure of the decision context. Assuming the bias inflates the perceived likelihood of the middle partition, the directional effect of such inflation on risk taking depends on whether optimal choice, under one's preferred benchmark, predicts the low or high risk option. As we discuss in a subsequent section, the heuristic may systematically favor increased risk taking in other contexts. More generally, the reduction in disparity across the value of options under partition bias, particularly when coupled with an allowance for decision noise, also helps explain why the heuristic predicts greater heterogeneity in risk than EU benchmarks. Predicted heterogeneity under heuristic choice increases further if one were to assume variation in bias severity or heuristic adoption.

Motivating Evidence for PPD Heuristic. The PPD heuristic departs from standard models of decision-making through three non-standard assumptions—pairwise evaluation, partition-dependent beliefs in the context of pairwise evaluation, and an allowance for decision noise—each supported by an extensive interdisciplinary literature. First, the tendency to evaluate stimuli through relative comparison is a well-documented psychological principle with evidence from neuroscience, decision science, psychology, and economics. As an example of the latter, the concept is integral to economic theories of reference-dependent preferences (e.g., Kahneman and Tversky 1979; Tversky and Kahneman 1992; Koszegi and Rabin 2007), salience and attention (e.g., Bordalo, Gennaioli, and Shleifer 2012, 2013), and comparative evaluation (e.g., Koszegi and Szeidl 2013; Bushong, Rabin, and Schwartzstein, 2021).

Second, the possibility that judgment and decision-making could be biased by the (potentially arbitrary) categorization imposed by the decision context was advanced by Fox and Rottenstreich (2003) and Fox and Clemen (2005). The theory, as applied to inference, builds on support theory, conceptualized by Tversky and Koehler (1994). Support theory posits that subjective perception of uncertain events may depend on the specificity with which such events are described, a phenomenon that could lead to violations of extensionality.[31] The influence of partition dependence has been documented across diverse, primarily experimental, contexts (see Benjamin, 2019). Of direct relevance to economics, Ahn and Ergin (2010) demonstrate how to incorporate partition-dependent beliefs into an axiomatic decision-theoretic expected utility framework. While intuitive application of partition dependence to three-option menus, such as those in GQ, would entail four partitions (i.e., performance less than goal 1, between goals 1 and 2, between goals 2 and 3, or greater than goal 3), our heuristic presumes that decision-relevant partitions emerge in the context of pairwise comparisons. By specifying the existence of three partitions, the heuristic arguably offers a more tractable framework for applying partition-dependent inference to (applicable) decision settings than contemplated by the literature. As a descriptive theory, it is also worth noting that partition-dependent beliefs are consistent with non-standard decision-making processes that have been recently advanced by economists (a point emphasized by Ahn and Ergin, 2010).[32]

A third assumption of the proposed heuristic is an allowance for decision error. The allowance of decision noise or error in judgment is a common feature of decision-making models within and outside of economics; its importance for choice and judgment has been emphasized in recent work such as Kahneman et al. (2021). Decision noise could arise from mechanisms such as bounded rationality, stochastic preferences, or limited cognition.

6.4 Experimental Evidence for the PPD Heuristic

We employ two strategies to assess the plausibility of the proposed heuristic for explaining risk taking. First, we present evidence from a new experiment (Experiment C) designed to test whether individuals adopt the key process assumptions underlying the heuristic and whether menus intended to discourage or encourage partition dependence generate variation in choice efficiency relative to baseline menus. Second, we assess whether the proposed heuristic explains a greater share of employee choice in the lab and the field than prior benchmarks.

Overview. We administered Experiment C in July 2022 on the Qualtrics platform to 893 employed US adults recruited from Amazon Mechanical Turk. After describing the real-life GQ paradigm, we randomized participants who successfully completed a comprehension check to one of two experimental arms. Both arms asked participants to make a hypothetical decision from a GQ menu identical to that featured in Experiment B (goals: 105 units, 110 units, 115 units; rewards: $150, $450,

[31] Fox and Clemen (2005) discuss how partition dependence also derives from the pruning bias (Fischhoff et al., 1978).
[32] Such mechanisms include the neglect or misunderstanding of contingencies (Martínez-Marquina et al., 2019; Sunstein and Zeckhauser 2010; Sunstein, 2002), noisy or imperfect memory (see Hilbert, 2012), or selective attention and informational salience (e.g., Bordalo, Gennaioli, and Shleifer 2012; 2013). For tractability, we follow Fox and Clemen (2005) in modelling the bias as a convex combination of partition-independent beliefs and a naïve prior that evenly distributes probability across partitions (a representation consistent with insufficient adjustment from a naïve prior). Our analyses do not preclude other representations of the bias.

$900). A first arm was designed to test whether participants adopted proximal pairwise comparisons, whether pairwise comparisons exhibited inferential bias consistent with partition dependence, and whether the degree of inferential bias, conditioned on partition-independent beliefs, predicted goal choice. A second arm was designed to test whether we would observe variation in choice efficiency across menus whose framing either encouraged or discouraged partition dependence.

Specifically, participants in the first arm were provided a fictional distribution of prior sales figures (reflecting actual menu-specific averages from the field) to increase the realism of their goal choice. After goal selection, we asked participants to introspect as to how they arrived at their decision by indicating which—if any—sets of two or three goals they directly compared. We then asked them to estimate their perceived likelihood of attaining each goal through both non-contingent and contingent elicitations (as an example of the latter, we asked participants to estimate the likelihood of reaching Goal 3 given certain knowledge of reaching Goal 2), reasoning that the contingent elicitations would be subject to the same pairwise partition bias hypothesized by the heuristic. Finally, to generate between-subject evidence for pairwise inferential bias and to test the generalizability of the phenomenon, we elicited contingent and non-contingent weather forecasts.[33]

Participants in the second arm were randomized to one of three menus that, while featuring the same goals and rewards, varied the framing of likelihood information. The first menu (*baseline*) communicated that a participant's likelihood of goal attainment was 83 (Goal 1), 74 (Goal 2), and 65 (Goal 3) percent (e.g., "You have an 83 percent chance of achieving Goal 1"), figures reflecting empirical averages from the field. A second menu (*partition independent*) displayed the identical likelihood for Goal 1 but contingently displayed the likelihood of Goal 3 given Goal 2 attainment implied by baseline beliefs ("If you achieve Goal 1, you have an 89 percent chance of also achieving Goal 2") and for Goal 3 ("If you achieve Goal 2, you have an 88 percent chance of also achieving Goal 3"). We hypothesized that this frame would lead to more accurate, and less partition dependent, inference than baseline despite their information equivalence. Finally, a third menu (*partition dependent*) displayed the non-contingent likelihood for Goal 1 and contingent likelihoods for Goals 2 and 3 modified to reflect moderate partition bias (we replaced the 89 and 88 percent from the prior condition with 74 and 65 percent, respectively).[34]

---

[33] We randomized participants to either forecast the likelihood that tomorrow's high temperature would be at least 70, 80, and 90 degrees Fahrenheit or to forecast the conditional likelihood of at least 90 degrees given certain knowledge of at least 80 degrees. Due to the difficulty of eliciting contingent beliefs, we piloted different communication strategies before adopting the following wording: "Suppose that you have a time-travelling friend who travels into the future. The friend returns and truthfully tells you that tomorrow's high temperature will be at least 80°F. Knowing for certain that the high temperature tomorrow will be at least 80°F, what are the chances that tomorrow's high will be at least 90°F?". To increase statistical power, this question was included in both experimental arms.

[34] Defining bias as a convex combination between non-contingent beliefs and a naïve prior, the contingent probabilities displayed in the biased contingent condition imply a $\theta = .44$ (comparison of Goals 1 and 2) and $\theta = .63$ (comparison of Goals 2 and 3). These parameter choices reflect an earlier representation of partition bias as the neglect of non-focal contingencies.

We hypothesized that this frame would reaffirm the partition bias naturally invoked from heuristic engagement of the baseline menu.

Results. The experiment yielded evidence consistent with widespread adoption of the proposed heuristic. First, regarding the process assumptions underlying the heuristic—86% of participants reported using pairwise comparisons in their deliberations; 93% of such participants made at least one proximal comparison. Participants also substantially underestimated conditional pairwise likelihoods relative to the Bayesian likelihood implied by non-contingent elicitations. That is, participants underestimated the likelihood of Goal 2 | Goal 1 attainment by 23% (0.64 relative to 0.82) and underestimated the likelihood of Goal 2 | Goal 1 attainment by 22% (0.59 relative to 0.76)—consistent with a partition bias parameter of $\theta = 1.01$ (Goals 1 and 2) and $\theta = 0.81$ (Goals 2 and 3). The elicitations of contingent and non-contingent weather forecasts implied an even more severe underestimation of 38 percent. Next, we found that the absolute magnitude of inferential bias associated with goal attainment strongly predicted efficient goal choice (as defined by a risk neutral EU benchmark), even after controlling for non-contingent beliefs. The regression estimates imply that eliminating the inferential bias associated with Goal 3 attainment would produce a 37% increase in the share of optimal choice (from 0.37 to 0.51).[35]

Last, we document a marked increase in EV/EU maximizing choice from menus designed to discourage partition dependence. Specifically, 61% of participants in the *partition independent* condition selected the EV-optimal goal, a 48% increase relative to the informationally-equivalent baseline (from 0.41 to 0.61, $p = 0.002$). Participants who engaged the *partition independent* menu also chose the EV-minimizing goal with significantly less frequency than baseline. As further evidence for heuristic choice, the 39% share of optimal choice associated with the partition dependent menu was not statistically distinguishable from either the baseline menu ($p = 0.82$) or the menu from the first experimental arm with no explicit display of attainment likelihoods ($p = 0.55$).

Pairwise versus Global Partition Bias - Experiment A. As noted earlier, a more straightforward application of partition dependence to GQ menus would be to assume inferential bias with respect to the full set of four partitions demarcated by the three goals rather than the three partitions associated with pairwise comparisons. We appeal to Experiment A, which elicited choice from menus of varying length, for evidence as to the appropriate implementation of partition bias. The experiment offers little support

---

[35] We estimated an additively linear model of each participant's EV-optimal goal choice, g^*, as a function of non-contingent beliefs of goal *k* attainment, $\hat{s}_k$, and the absolute bias in relative inference implied by contingent beliefs, $\hat{\lambda}_{k,k-1} = |\hat{s}_k/\hat{s}_{k-1} - \hat{s}_{k|k-1}|$: $g^* = \alpha + \pi_1\hat{s}_1 + \pi_2\hat{s}_2 + \pi_3\hat{s}_3 + \gamma_1\hat{\lambda}_{3,2} + \gamma_2\hat{\lambda}_{2,1} + \varepsilon$. The regression (excluding six observations with non-unique optima) indicates that while the perceived likelihood of attaining Goal 3 (the EV-optimal goal for most participants based on their non-contingent beliefs) strongly, and expectedly, predicts EV-optimal choice ($\hat{\pi}_3 = 1.00$, $p < 0.001$), the magnitude of inferential bias strongly (negatively) predicts choice ($\hat{\gamma}_1 = -0.81$, $p < 0.001$). The substantial partial correlation between inferential bias and optimal goal choice was robust to a variety of alternative non-parametric specifications.

for a global implementation of the bias. For example, given θ = 0.75 (a degree of bias roughly consistent with, but slightly more conservative than, Experiment C), global implementation accurately predicts the full set of decisions for 0% of participants (10% allowing for decision noise)—rates far less than pairwise application of the bias, as we discuss in the subsequent section.

6.5 Predictive Accuracy of the PPD Heuristic – Lab and Field

Perhaps the most instructive test of the pairwise heuristic is to gauge the accuracy with which it explains observed behavior. To assess its explanatory power in the lab (Experiment A and first arm of Experiment C) and field and to evaluate the relevance of constituent assumptions, Table 7 compares predictive accuracy of heuristic choice across a range of specifications to a risk neutral SEU baseline. The table specifically considers heuristic choice across varying representations of bias—no bias, parameterized bias of θ = 0.75, and the experimentally estimated person-specific bias (Experiment C only)—and +/- noise allowances intended to subsume plausible levels of decision imprecision.[36]

Across domains, the table documents substantially higher predictive accuracy of heuristic choice than under baseline—or previously tested—benchmarks. The heightened accuracy appears attributable both to the assumption of inferential bias and the inclusion of modest decision noise. In the field, assuming a noise allowance of either $25 or $50, the parameterized heuristic explains 83 to 92% of employee choice, a 66 to 84% improvement over baseline. Analyses of Experiment C, where we can compare predictive accuracy of the heuristic, assuming the same noise allowances, under the personalized (59 to 72%) and parameterized (49 to 63%) bias, suggests that estimates of accuracy under the parameterized bias in the field may understate the true explanatory power of the heuristic, ignoring possible ceiling effects. Finally, the improvement in predictive accuracy from Experiment A, where we observe multiple decisions per participant and a low rate of baseline accuracy, far exceeds that observed in the field on a relative basis (8.5 to 13.8x) but is roughly comparable in absolute terms.

Revisiting Gender Difference in Risk Taking. The PPD heuristic also offers an intriguing potential explanation for previously discussed gender difference in conservative choice. The heuristic (θ = 0.75; ω = 25) implies a gender difference in conservatism of 0.04—a gender gap 64 to 71% smaller than the gaps implied by standard EU benchmarks (0.11 to 0.14). While gender gap reduction is less pronounced on a relative basis, this owes to the substantially lower overall rate of risk aversion implied by the heuristic. For additional insight as to how heuristic choice might help to explain gender differences in choice, we return to Experiment C. The experiment reveals only small, insignificant, differences in the

[36] The exercise tags as accurate any choice that maximizes subjective EV subject to the noise allowance. To provide context, a noise of allowance of $25 is equivalent to 10% of the EV difference between Goals 2 and 3 and 12% of the EV difference between Goals 1 and 2 from the GQ-representative menu featured in Experiment C. The noise allowances for Experiment A are roughly analogous to those used for Experiment C on a proportional basis.

magnitude of inferential bias across gender, either in the context of goals or weather. One interpretation of these patterns is that the observed gender gap in conservative choice may reflect the higher frequency with which women apply the heuristic relative to men. Such an interpretation highlights the possibility that gender (or other sub-group) differences in risk taking may reflect heterogeneity in decision strategies, such as heuristic adoption, rather than systematic differences in risk preferences and/or perceptions.

6.6 Explaining Residual Goal Choice – Narrow Application of Heuristic

We interpret evidence from the experiment and predictive analyses as suggesting that a moderate to large share of DMs relied on a decision strategy resembling the proposed heuristic. One approach for estimating the lower bound of heuristic adoption is to calculate the difference in the share of explained choices between the standard benchmark and the heuristic. This approach, using field data (assuming noise allowances of either $25 or $50), yields a lower adoption bound between 33 and 42 percent. Given many employee choices consistent with standard benchmark predictions are also consistent with heuristic predictions, the table implies far higher upper bounds of adoption. While we cannot precisely estimate the adoption rate, it is notable that the heuristic explains a higher share of choice also explained by the risk neutral SEU benchmark than the converse. For example, in the field, assuming a noise allowance of $25, the heuristic explains 96% of choice also explained by the benchmark and 69% of choice unexplained by the benchmark. Conversely, the benchmark explains only 58% of choice also explained by the heuristic and 11% of choice unexplained by the heuristic.

The decisions of some employees do not adhere to any of the benchmark models we consider, including the proposed heuristic. While some of these decisions likely reflect confusion, inattention, or idiosyncratic strategies, we speculate that some employees may have narrowly applied a decision rule to a subset of the menu. For example, GQ employees may have only concretely evaluated Goals 1 and 2, having ruled out Goal 3 for unknown reasons. Returning once again to Experiment C for additional insight, we find that roughly one-quarter of participants whose choices were inconsistent with the heuristic reported evaluating only Goals 1 and 2—an indication of process consistent with their final goal choices. Applying heuristic choice to only the first two goals explained all but one of these decisions.

7 APPLYING PPD HEURISTIC TO INSURANCE PLAN CHOICE

We speculate that the proposed heuristic could offer insights into financial risk taking extending beyond employee decisions in reward programs.[37] Specifically, we see the heuristic as applicable to economic contexts where DMs engage a menu of risky options that can be conceptualized as a choice

[37] Employee reward programs such as GQ are arguably of independent economic interest given their popularity. GQ estimates that roughly 40 percent of Fortune 500 firms have adopted their program since its 2001 inception.

between nested lotteries of ascending risk and reward (i.e., lotteries that draw from a common source of risk such that the set of winning outcomes of a less risky option entirely subsume that of a more risky one). One can reasonably interpret economic menus involving contingent labor contracts, gambling, (retirement) portfolio allocation, options investing, and insurance plan choice in this manner. For example, consider the decision to allocate one's retirement savings via an ordered menu of risk-varying fund options (e.g., from fixed income products to small cap/growth equity). Heuristic choice could lead to sub-optimal risk aversion if investors systematically underestimated the relative likelihood of high market returns in the context of pairwise fund comparisons. While comprehensive discussion of each of these domains is beyond the present scope, in this section we discuss the particularly promising application of the heuristic to consumer insurance markets in greater detail.

7.1 Theoretical Framework for Heuristic Plan Choice

Consumer insurance offers a natural analogue to GQ since insurance menus often feature plans that vary only in their cost and cost-sharing (one distinction, of course, is that insurance plans vary in non-contingent costs or annual premia, whereas GQ participation is costless).[38] We explore the applicability of heuristic plan choice to insurance demand more formally by outlining a theoretical decision framework, deriving the market conditions under which heuristic choice leads to systematic bias and excess heterogeneity in demand, and administrating a series of experiments to test whether heuristic choice can help to resolve seemingly contradictory puzzles from the empirical literature.

<u>Model Notation and Setup</u>. We begin by considering the stylized decision of a risk-neutral, utility-maximizing, DM tasked with purchasing insurance to protect against the prospect of random loss, $x \geq 0$.[39] The DM selects from a menu of two plans, j ϵ $[l, h]$, varying only in annual premia, $p_h > p_l$, and cost-sharing, $b_h(x) \geq b_l(x)$, where $b_j(.)$ is a function that specifies an indemnity payoff for a given loss. Both plans cover losses beyond some out-of-pocket maximum such that plan value, or utility, can be described by $u_j(x) = b(x) - x - p_j$. For any potential loss, we denote the difference between the high and low coverage plans in expected utility by Δu_s, indemnity payoff by Δb_s, and premia by Δp.

As before, we stipulate that the DM evaluates the plans via pairwise comparison. Such relative comparison causes the DM to partition potential loss outcomes into three decision-relevant states. A first state, $S_L \equiv x < X_L$, specifies realizations of loss for which neither plan provides coverage. For example, given two plans whose cost-sharing takes the exclusive form of a deductible, $D_h < D_l$, the first partition threshold would be the deductible of the high coverage plan, $X_L = D_h$, since losses below this threshold

[38] In practice, we believe the analogy holds if one can credibly price non-financial differences across plans and cost-sharing is primarily achieved through a single cost-sharing channel such as a deductible or coinsurance.

[39] While the assumption of risk neutrality departs from typical treatment of insurance choice, in contexts where plans differ sufficiently in their expected value, the assumption simplifies analysis without sacrificing generality.

would yield no indemnity payoffs for either plan while losses immediately above would yield payoffs for only the high coverage plan. As another example, for plans whose cost-sharing differences are reflected in different coinsurance rates, $X_L = 0$. A second state, $S_H \equiv x > X_H$, specifies realizations of loss for which the high coverage plan provides sufficient additional coverage, relative to the low coverage plan, to warrant its higher premium (i.e., $\Delta b_h > \Delta p$ or $\Delta u_h > 0$). Finally, a third state, $S_M \equiv X_L \leq x \leq X_H$, describes losses between the two thresholds. The perceived likelihood of state S_K is denoted by $\widehat{\varphi}_K$.

Pairwise Decision Rule. A risk-neutral, utility-maximizing, DM, will enroll in the high coverage plan if: $\widehat{\varphi}_{L+}[\widehat{\varphi}_{H|L+}\Delta b_H + \widehat{\varphi}_{M|L+}\Delta b_M] > \Delta p$. The expression, written to reflect the presumed propensity of DMs towards relative comparison, simplifies to: $\widehat{\varphi}_H\Delta b_H + \widehat{\varphi}_M\Delta b_M > \Delta p$. The rule implies that a DM will opt for greater coverage if the expected gain from such coverage in the event of severe, Δb_H, or moderate, Δb_M, loss outweighs the difference in plan costs, Δp. Absent inferential bias, $\widehat{\varphi}_K = \varphi_K$, and decision noise, this rule predicts utility-maximizing choice. Given heuristic choice with pairwise partition dependence of the form, $\widehat{\varphi}_K = \theta/\pi + (1 - \theta)\varphi_K$ and $\pi = 3$, choice depends not only on plan differences in cost and cost-sharing and DM expectations of loss risk, but also the degree of bias, θ, and the economic structure of the market, roughly proxied by $\varphi_H + \varphi_M$ (going forward, we ignore decision noise for simplicity). As previously discussed, one can adapt the heuristic to ordered menus with more than two options by specifying an initial pairwise comparison and stopping rule. For example, one could assume consumers successively compare proximal pairs of plans beginning with the lowest coverage option and stopping their evaluation once a higher coverage plan is rejected. In large menus where consideration of all options is impractical, it is reasonable to expect DMs to apply heuristic choice to a focal subset of options.

Bias and Heterogeneity in Insurance Demand. To better understand the direction and magnitude of bias implied by heuristic choice across insurance markets, we can derive net bias as a function of model parameters. After differencing decision rules with and without bias, rearranging terms, and denoting bias with respect to φ_K as $\tau_K = \widehat{\varphi}_K - \varphi_K$, we arrive at the following expression for abnormal willingness to pay for increased coverage, in terms of Δp: $\tau_H\Delta b_H + \tau_M\Delta b_M$. The sign of the expression indicates the direction of bias in market demand predicted by the heuristic. Figure 7 provides graphical intuition for the expression by plotting demand bias under heuristic choice for varying market structures, menu configurations, and bias severity. The first panel conveys the negative relationship between bias and a market's baseline loss risk and how bias severity amplifies this relationship (fixing $\widehat{\varphi}_M/\widehat{\varphi}_H = 1/2$). That is, heuristic choice predicts excess demand, relative to standard benchmarks, in markets characterized by relatively low perceived loss risk (e.g., home, vehicular, health insurance) and excessively low demand in markets characterized by high perceived loss risk (e.g., dental insurance, vision care, prescription drug coverage).[40] The second panel of the figure conveys how menu configuration can interact with market structure to shape demand bias (fixing $\theta = 0.75$). For example, given increases in plan cost sharing or in the number of available plan options—both likely implying a fall in $\hat{\varphi}_M/\hat{\varphi}_H$—the sensitivity of heuristic bias to market structure rises (due to a steepening of the bias gradient).

We interpret heuristic choice as predicting greater heterogeneity in risk taking than predicted by standard benchmarks across most markets. One reason for increased heterogeneity is that the reduced disparity across plans in perceived value under heuristic choice should lead to greater choice diversity, particularly when allowing for decision noise. In addition, any heterogeneity in the severity of inferential bias, or in the adoption of the heuristic, would predict additional choice diversity.

7.2 Experimental Evidence on Heuristic Plan Choice

A large literature in economics has investigated the motives for consumer insurance demand across a variety of markets and has documented repeated instances in which utility-based risk preferences cannot account for the level and/or heterogeneity of demand (see Barseghyan et al. 2018). We conclude with a series of experiments designed to test whether heuristic choice can help to explain two specific empirical puzzles, with opposite valence, from this literature. The first puzzle involves the propensity of many high-risk elderly Medicare Part D enrollees to undersubscribe, relative to standard benchmark predictions, high-coverage plans (e.g., Abaluck and Gruber, 2011; Heiss et al., 2013). Such a situation corresponds to an insurance market with a high baseline likelihood of at least some covered loss—for most pairwise comparisons, practically implying a high $\hat{\varphi}_M + \hat{\varphi}_H$. In this scenario, the heuristic predicts under-insurance, relative to standard benchmarks, due to systematic underestimation of high loss risk. The second puzzle involves the propensity to over-insure in the market for home insurance—a market typically characterized by low baseline loss risk (e.g., low-deductible enrollees had an average claim rate of less than 5% in the sample examined by Sydnor, 2010). In this scenario, the heuristic predicts excess insurance demand due to systematic overestimation of moderate and high loss risk.

<u>Prescription Drug Coverage</u>. A first experiment (Experiment D) investigated whether heuristic choice can help explain sub-optimally low demand for prescription drug coverage by examining the sensitivity of hypothetical plan choice to variation in the framing of loss risk. Specifically, we asked 432 US adults, recruited from Amazon Mechanical Turk, to imagine they were about to enter retirement and had to decide whether to purchase prescription drug insurance (participants were informed they already had separate coverage for non-prescription medical expenses). After an educational module explained

[40]A potential exception to this taxonomy is catastrophic insurance, a setting long characterized by sup-optimally low insurance demand (e.g., Kunreuther 1973). While the literature has sought to explain low demand through non-standard decision motives and market failures, many of these explanations involve consumers who do not actively engage plan choice. We speculate that in catastrophic care settings with mandatory coverage requirements, the heuristic would lead to excess insurance demand.

how drug bills mapped to out-of-pocket costs under various cost-sharing scenarios, participants were asked to select from a menu of two plan options (Silver, Gold) varying only in annual premia ($640, $1220) and cost-sharing (coinsurance rates: 50%, 15%). They were also given the option of selecting no plan. The menu resembled Medicare Part D in that plans varied in cost-sharing primarily via coinsurance, covered all expenses beyond a fixed out-of-pocket threshold ($7,500), and had medal-color labels.

To facilitate plan choice, we provided participants with plausible information on the distribution of prospective drug bills costs for a real-life high-risk Medicare Part D enrollee. We communicated the likelihood that drug bills would exceed three decision-relevant thresholds ($0, $1,280, $1,657), truthfully explaining that if drug bills exceeded the second threshold, the Silver or Gold Plan would minimize total costs and if they exceeded the third threshold, the Gold Plan would minimize costs. For tractability, we additionally conveyed that drug bills would never exceed $10,000 in a year—even for those selecting no plan—and that they would follow a uniform distribution between conveyed thresholds. We assigned participants to one of two sets of likelihood thresholds (80%, 60%, 40%; 80%, 63%, 48%) deemed plausible and roughly resembling thresholds from earlier GQ experiments. Given the provided information, the Gold Plan comfortably minimized expected total annual cost (premium + out-of-pocket).

While all participants engaged the same set of plans, they were randomized to one of three experimentally varying menus. The first menu displayed cost information with respect to the thresholds in non-contingent terms (*baseline*) (e.g., “You have a 60% chance of a drug bill exceeding $1,280”). The second menu displayed cost information contingently after adjustment to reflect moderate to severe partition dependence with respect to the comparison between the Gold and Silver Plans (*partition dependent*) (e.g., “If your drug bill exceeds $1,280, you have a 40% chance of a drug bill exceeding $1,657”).[41] The third menu, intended to discourage partition dependence, displayed contingent costs without adjustment for bias (*partition independent*) (e.g., “If your drug bill exceeds $1,280, you have a 67% chance of a drug bill exceeding $1,657”).

The outcome of the experiment, summarized in Appendix Table A5, produced three patterns of note. First, consistent with the literature, baseline participants exhibited substantial demand for non-EV-optimal plans, with only 30% selecting the cost-minimizing Gold Plan. Second, the baseline share of EV-optimal plan choice was nearly identical to that produced by the *partition dependent* menu which displayed contingent, but biased, cost information (29%; $p = 0.82$).[42] Third, relative to baseline, participants chose far more optimally (39%) from the *partition independent* menu ($b = 0.12$; $p = 0.03$).

[41] The probabilities displayed in the partition dependent condition reflect an earlier implementation of partition dependence that assumed full neglect of the non-focal, low-state, partition, s_L. In the context of the present model, the displayed probabilities correspond to $\theta = 0.64$ (No Plan compared to Silver Plan) and $\theta = 1.71$ (Silver Plan compared to Gold Plan) for the first set of likelihoods (80%, 60%, 40%) and, analogously, $\theta = 0.59$ and $\theta = 1.08$ for the second set of likelihoods (80%, 63%, 48%).
[42] Treatment effect estimates generated from LPM regressions of the specified outcome on treatment and menu indicators.

Overall, the experiment revealed significant improvements in choice efficiency across menus intended to discourage partition bias, as predicted by heuristic choice. As with actual Medicare Part D, the economic consequences of inefficient choice were meaningful— selecting the Silver Plan instead of the Gold Plan implied $452 in additional annual costs in expectation, equivalent to 37% of the Gold Plan premium.

Home Insurance. A second experiment (Experiment E) investigated whether the proposed heuristic could help explain the over-insurance described in analyses of the US home insurance market. In slight contrast to the previous study, the experiment compared hypothetical plan choice across menus strategically varying both the presence and framing of potential loss information to either encourage or discourage partition bias. Given the low baseline risk of loss, we hypothesized that increasing the salience of the no-loss likelihood would diminish the relevant partition bias, increasing choice efficiency. We implemented the study by asking 435 US adults, recruited from Amazon Mechanical Turk, to select lender-mandated insurance from a menu of three plan options adapted from Sydnor (2010). The options varied only in annual premium and deductible: (1) Basic Plan ($1,000 deductible, $616 premium) (2) Medium Plan ($500 deductible, $716 premium), (3) Premium Plan: ($250 deductible, $803 premium).

We randomized participants to one of four experimentally varying menus. An initial baseline menu (*full information baseline*) accurately conveyed potential loss information roughly adapted from Sydnor (2010): a 4% likelihood of any loss; conditioned on some loss, a 75% likelihood of severe loss ($2500+) and a 25% likelihood of non-severe loss ($1 to $2500). After simplifying cost assumptions, the specified potential cost distribution implied the EV-optimality of the Basic Plan.[43] A second menu (*no information baseline*) conveyed no information regarding potential costs. A third menu (*partition dependent*), designed to encourage partition dependence, displayed the 75% conditional likelihood of a loss but did not display the likelihood of any loss. A fourth menu (*partition independent*), designed to discourage partition dependence, conveyed the 96% likelihood of no loss—a framing deemed to be more salient than the 4% likelihood of any loss—without further displaying the relative likelihood of loss severity. Despite the varying presence of actual risk information across conditions, we interpret the *full information baseline* and *partition independent* menus as effectively equivalent economically since no partitioning of severe and non-severe loss, given the stated likelihood of any loss, could rationalize EV-optimal demand for anything but the Basic Plan.[44]

---

[43] Expected cost calculations assume a 3% likelihood of severe losses with $2500 in damages and a 1% likelihood of non-severe losses with $500 in damages. The rank order of plans by expected value—Basic ($626), Medium ($707), Premium: ($781)—is not sensitive to the assumed distribution of costs within the specified range.

[44] While improbably high perceptions of baseline risk could rationalize demand for costlier plans in the no information baseline and focal contingency conditions, in conditions conveying baseline risk, the most pessimistic beliefs about relative severity can only rationalize the choice of the Basic Plan by an EV-maximizing consumer.

The experiment, summarized in Appendix Table A6, once again, corroborated findings from the empirical literature and heuristic predictions of the heuristic. First, across both baseline conditions, most participants chose a non-EV-maximizing plan, paralleling the over-insurance observed in the real-life home insurance markets. Second, when encouraged to actively attend to the high no-loss likelihood, participants demanded substantially less insurance than in either baseline condition—i.e., participants in the *partition independent* condition were 35 (54) percent more likely to select the EV-maximizing plan and 58 (69) percent less likely to select the EV-minimizing plan compared to the *no information* (*full-information*) *baseline*. Third, participants demanded greater coverage from the *partition dependent* menu than in any other condition, consistent with participants overestimating baseline loss risk under heuristic choice. To rationalize plan choice, under the EV benchmark, the 77% of participants not selecting the Basic Plan would have had to assume an implausibly high annual loss rate of 27+ percent (the possibility that these choice patterns simply reflect sharply inflated perceptions of baseline risk is not supported by participants demanding less insurance from baseline menus with no information versus full information.

While Experiment E offers evidence consistent with heuristic predictions for a specific low level of baseline loss risk, the heuristic also predicts the decay of over-insurance bias across markets with increasing baseline risk (Figure 7). We sought to test this prediction through a final experiment involving home insurance choice (Experiment F). Specifically, we asked 348 US adults, recruited from Amazon Mechanical Turk, to select insurance from menus featuring the same plans as the prior experiment. To credibly vary the likelihood of baseline loss risk without having to shift markets, we randomized participants to one of three conditions varying the stated duration of plan coverage from one year, five years, and ten years (e.g., for the latter conditions, we explained that "The plan is different than most in that it provides coverage for the next [5, 10] years [bolded] with a single deductible and one-time premium."). Within each of these conditions, participants were cross-randomized to either a *baseline* menu displaying no loss information or a *partition independent* menu displaying the cumulative likelihood of no loss (e.g., "To help you make a decision, suppose that during your plan period, there's a roughly 77 percent [bolded] chance of <u>no damage</u> to your home."). The randomization resulted in a 3 (baseline risk) x 2 (risk salience) between-subject design. Assuming an annual loss likelihood of 5% and earlier simplifying cost assumptions, participants would have minimized total expected costs under the Basic Plan given a one year duration (5% loss risk), the Medium Plan given a five year duration (23% loss risk), and the Premium Plan given a ten year duration (40% loss risk).

As depicted in Figure 8, the experiment yielded choice patterns consistent with the heuristic. First, participants exhibited greater over-insurance with baseline loss risk of 0.05 compared to 0.23 ($p < 0.001$) (over-insurance was not possible with loss risk of 0.40). Second, participants in the *partition independent* condition over-insured at a markedly lower rate than at baseline ($p = 0.006$). Finally, while the interaction was not statistically significant, *partition independent* participants were more likely to choose optimally than baseline in the lowest loss-risk setting but less likely to choose optimally in the highest loss-risk setting. Collectively, these patterns cannot be explained through random choice, utility-based risk preferences, or overestimation of baseline risk. Beyond suggesting the importance of inferential error in insurance evaluation, the experiments allude to multiple violations of descriptive invariance, an axiomatic assumption of most economic analyses of insurance. Such violations highlight the opportunity to systematically improve choice efficiency through strategically reframed menus.

8 CONCLUSION

We present new insights as to the prevalence of financial risk-taking and its underlying motives. Our main evidence describes the decisions of several thousand employees in the context of a popular all-or-nothing employee goal-reward program. In contrast to other field settings, we see this setting as uniquely helpful for understanding risky choice given the simple and standardized decision context, the diversity of employees and economic stakes, the near-complete participation rate, and our ability to observe contemporaneous employee perceptions of risk. A central finding is to document substantial risk aversion and choice heterogeneity in employee decisions, resulting in significant counterfactual reward loss. The excess conservativism of employees was robust to reward size and employee tenure and was substantially higher for women than men. In investigating risk-taking motives, we document how these choice patterns cannot be explained through plausible utility-based risk preferences, even allowing for preference heterogeneity, nor can they be explained via prominent non-standard explanations from the literature such as biased beliefs, non-linear decision weights, or gain-loss utility. We corroborate these choice patterns—and the limited explanatory power of previously tested benchmark models—via an incentive-compatible, online goal-reward paradigm with verified comprehension, dollar-denominated rewards, and limited scope for signaling, reputation, or effort costs. We further corroborate patterns from a menu explicitly recasting goal options as nested lotteries with known probabilities.

We propose a novel heuristic explanation for conservative and heterogeneous choice involving simple departures from the EU framework for which there is considerable support in the literature. The heuristic stipulates that employees select goals through a series of approximate and proximal pairwise comparisons. Due to partition-dependent inference, the heuristic posits these pairwise comparisons will lead many employees to substantially underestimate the relative likelihood of risky outcomes, resulting in more conservative (and heterogeneous) choice than standard benchmark predictions. After validating the heuristic experimentally, we show that it explains a greater share of choice in the lab and in the field than previously considered benchmarks. And by accounting for most of the gender gap in conservative choice, the heuristic further suggests that apparent gender differences in risk taking may reflect differences in decision strategy rather than systematic differences in risk preferences or beliefs.

We speculate that the proposed heuristic may help understand a broad range of risk taking in contexts where choice options can be conceptualized as an ordered set of nested lotteries. Towards this end, we apply the heuristic to insurance plan choice and illustrate how it predicts systematic bias in demand—relative to standard benchmarks—of a direction and magnitude determined by structural features of the market. A last set of experiments demonstrate how the heuristic offers a potential explanation for seemingly contradictory puzzles from the empirical literature on consumer insurance demand in addition to providing an explanation for the unexplained heterogeneity routinely cited in the same literature. We hope future work will clarify the specific cognitive processes underlying the proposed heuristic, explore its applicability to other risk taking domains, and work out its implications for welfare analyses and policy design in markets such as insurance.

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.

Baillon, A., Bleichrodt, H., & V. Spinu. (2020). Searching for the Reference Point. *Management Science*, 66(1): 93–112.

Bandiera, O., Parekh, N., Petrongolo, B. & Rao, M. (2022), Men are from Mars, and Women Too: A Bayesian Meta-analysis of Overconfidence Experiments. *Economica*, 89: S38-S70.

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): 501–64.

Benartzi, S., & R. Thaler. (2007). Heuristics and Biases in Retirement Savings Behavior (2007). *Journal of Economic Perspectives*, 21(3): 81–104.

Benjamin, Daniel J. (2019). Chapter 2 - Errors in probabilistic reasoning and judgment biases. In B. Douglas Bernheim, Stefano DellaVigna, David Laibson (eds.), *Handbook of Behavioral Economics: Applications and Foundations 1, Volume 2* (pp. 69-186). 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., Gennaioli, N. & Shleifer, A., 2013. Salience and consumer choice. *Journal of Political Economy*, 121(5): 803-843.

Bushong, B., M. Rabin, & J. Schwartzstein. (2021). A model of relative thinking. *The Review of Economic Studies*, 88(1): 162–191.

Cohen, A. & L. Einav (2007). Estimating risk preferences from deductible choice. *American Economic Review*, 97(3): 745–788.

Cutler, D., & R. Zeckhauser. (2004). Extending the Theory to Meet the Practice of Insurance. In R. Litan & R. Herring (eds.), Brookings-Wharton Papers on Financial Service. Washington, D.C.: Brookings Institution.

Ericson, K., & A. Starc. (2012). Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts Connector. *The American Economic Review*, 102 (3): 493-97.
Fischhoff, B., P. Slovic, S. Lichtenstein. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception Performance*, 4(2): 330–344.
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., Erner, C., & D.J. Walters. (2015). Decision under risk: From the field to the laboratory and back. The Wiley Blackwell handbook of judgment and decision making, 2: 41-88.
Fox, C.R., & Y. Rottenstreich. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, 14(3): 195-200.
Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica*, 59(3): 667–686.
Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7): 2643–2682.
Handel, B. R. & J. Kolstad. (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8): 2449–2500.
Heath, C., Larrick, R. P., & G. Wu. (1999). Goals as Reference Points. *Cognitive Psychology*, 38(1): 79–109.
Heiss, F., Leive, A., McFadden, D., & Winter, J. (2013). Plan selection in Medicare Part D: evidence from administrative data. *Journal of Health Economics*, 32(6): 1325–1344.
Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, 138(2): 211–237.
Holt, C., & S. K. Laury. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92 (5): 1644-1655.
Jaspersen, J. G., Ragin, M. A., & 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.
Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise: a flaw in human judgment. First edition.
Kamenica, E. (2008). Contextual Inference in Markets: On the Informational Content of Product Lines. *The American Economic Review*, 98(5): 2127–2149.
Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1): 119-131.
Kőszegi, B., & M. Rabin. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121(4): 1133–65.
Kőszegi, B., & M. Rabin. (2007). “Reference-Dependent Risk Attitudes.” *The American Economic Review*, 97(4): 1047–73.
Kőszegi, B., & Szeidl, A., 2013. A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1): 53-104.
Kunreuther, H. (1973). Recovery from natural disasters; insurance or Federal aid? United States: American Enterprise Institute for Public Policy Research.
Kusev, P., Purser, H., Heilman, R., Cooke, A.J., Van Schaik, P., Baranova, V., Martin, R. & Ayton, P., 2017. Understanding risky behavior: The influence of cognitive, emotional and hormonal factors on decision-making under risk. *Frontiers in Psychology*, 8: 102.
Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92(368), 805–824.
Martínez-Marquina, A., Niederle, M., & E. Vespa. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109(10): 3437-74.
Niederle, M. & L. Vesterlund. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much?, *The Quarterly Journal of Economics*, 122(3): 1067-1101.

Niederle, M. (2017). Chapter 8. Gender. In J. Kagel & A. Roth (Ed.), The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics (pp. 481-562). Princeton: Princeton University Press.

Post, T., Assem, M., Baltussen, G., & R.H. Thaler. (2008). Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review*, 98(1): 38–71.

Prelec, Drazen. (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.

Rabin, M., & R.H. Thaler. (2001). Anomalies: Risk Aversion. *Journal of Economic Perspectives*, 15(1): 219-232.

Sugden, R. (2003). Reference-dependent subjective expected utility. Journal of Economic Theory, 111 (2), 172–191.

Sunstein, C.R. (2002). Probability Neglect: Emotions, Worst Cases, and Law. Yale Law Journal, 112, 61.

Sunstein, C.R., & R.J. Zeckhauser. (2010). “Dreadful possibilities, neglected probabilities.” The Irrational Economist. Making Decisions in a Dangerous World. Eds. Erwann Michel-Kerjan & Paul Slovic, Public Affairs Books, NY: New York.

Sydnor, J. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics*, 2(4): 177–99.

Tversky, A., & D. Kahneman. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5 (4): 297–323.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological review, 101(4), 547.

von Neumann, J., & O. Morgenstern. (1947). Theory of games and economic behavior (2nd ed.). Princeton, NJ: Princeton University Press.

APPENDIX

1. Characterizing Choice with CRRA Utility Benchmarks

Our primary analysis assessed employee goal choice for benchmark models featuring a utility function from the constant absolute risk aversion (CARA) family. The assumption of CARA utility, over the more common choice of constant relative risk aversion (CRRA) utility, was motivated by tractability given a lack of data on employee wealth. In this section we recharacterize choice for the core benchmark models assuming CRRA utility across a wide range of wealth and degrees of relative risk aversion.

Specifically, we assume employees are governed by CRRA utility of the form: $u(x) = \frac{x^{1-\rho}}{(1-\rho)}$ for $\rho \neq 1$ and $u(x) = \ln(x)$ for $\rho = 1$. We assess choice for initial lifetime wealth ranging from $1,000 to $1,000,000 and relative risk aversion, $\rho \in [0.10, 50]$. To appreciate the breadth of risk attitudes captured by the latter interval, we follow Post et al. (2008) in mapping risk parameters to the implied certainty coefficient—that is the certainty equivalent expressed as a fraction of expected value—associated with a 50/50 bet of ($0, $10k) assuming initial wealth of $25,000. This interval almost certainly subsumes the range of plausible relative risk aversion—asserted by Holt and Laury (2002) as bounded by 0 and 1.37.

Appendix Table A1 summarizes the choice characterization by reporting the optimal choice share for the EU benchmark across beliefs (rational, subjective), initial wealth, and relative risk aversion. The table indicates that within the (highlighted) range of plausible attitudes towards risk (spanning certainty coefficients from 0.87 to 0.99), the CRRA benchmarks explain a share of choice virtually identical to the CARA analogues from Table 4 assuming either risk neutrality or moderate risk aversion for rational expectations (0.45) and subjective beliefs (0.50). Overall, we interpret the table as suggesting that one cannot attribute the lack of descriptive accuracy of the benchmark models tested in the main analyses to the assumption of constant absolute, rather than constant relative, risk aversion.

2. Generalized Theoretical Framework with Discretionary Effort

While the theoretical framework from the main text abstracts away from goal choice motives pertaining to effort, here we present a more generalized baseline decision model incorporating discretionary effort. We then discuss the implications of the omission of effort for the characterizations estimated from the original framework. Specifically, we describe the decision of a risk neutral, utility-maximizing, employee who jointly selects a productivity goal from an all-or-nothing menu and some level of costly but productivity-enhancing effort. As before, we represent goal choice through lotteries, $G_n \in [G_h, G_l]$, yielding positive reward x_n, with some probability, s_n and 0 with some probability $(1 - s_n)$. Goals are associated with ascending productivity thresholds, such that attainment of G_h implies attainment of G_l. Employees must also commit to either low or high effort, $e \in [0, 1]$. The cost of effort is known to the employee, positive, and increasing in effort, such that, if we normalize low-effort costs to zero, we can denote high effort costs as $c > 0$. Higher effort weakly increases the likelihood of goal attainment, such that, $s_n(1) \geq s_n(0)$. The high goal has a strictly higher reward, $x_h > x_l$, and, for a given level of effort, a lower attainment likelihood, $s_h(.) < s_l(.)$, than the low goal. We assume an inter-temporal discount rate of 1, rendering the timing of reward receipt immaterial, and that effort commitments cannot be modified following goal choice.

Decision Rule. If $\hat{s}_n(e)$ denotes a risk neutral employee's perceived likelihood of goal attainment given some level of effort, the employee will choose a goal and effort level to maximize the following:

$$\max_{n\epsilon(h,l),e\epsilon(0,1)} \mathrm{U}(G_n, e) = x_n \hat{s}_n(e) - ec$$

Denoting optimal effort for the chosen goal as e^*, the alternative level of effort as e', and the difference in effort costs associated with a shift from e^* to e' as Δc, the employee will choose the high goal if the following two conditions are satisfied:

(1) Expected Utility Condition:

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

(2) Incremental Effort Condition:

$$x_h \hat{s}_h(e^*) > \hat{s}_l(e')x_l + \Delta c$$

Intuitively, the decision rule stipulates that for an employee to select a high goal, the favorability of its reward relative to the low goal, under optimal effort provision, must offset its lower likelihood. High goal choice additionally requires that its expected value, under optimal effort, exceeds low goal value under an alternative level of effort provision. If both conditions are not met, the rule specifies low goal choice.

Bounding Characterization Estimates. Our strategy for characterizing goal choice in the main text draws on observed goal choice, G_n, and, for each goal, beliefs of goal attainment assuming optimal effort under the chosen goal, $\hat{s}_n(e^*)$. It does not leverage beliefs of goal attainment for non-chosen goals reflecting optimal effort for such goals. Said differently, our characterization strategy relies on an ability to observe condition (1) but not condition (2). We can, however, derive how our estimates of optimal and conservative choice would directionally compare to counterfactual estimates under full information. We carry out this exercise by considering four possible characterization scenarios based on observed goal choice and satisfaction of condition (1):

i. Observe G_h and condition (1) satisfied: Our approach characterizes this scenario as optimal choice relative to benchmark predictions. It is possible, however, that if condition (2) is not satisfied, the scenario reflects aggressive choice.

ii. Observe G_h and condition (1) not satisfied: Our approach accurately characterizes this scenario as aggressive relative to benchmark predictions.

iii. Observe G_l and condition (1) satisfied: Our approach accurately characterizes this scenario as conservative relative to benchmark predictions.

iv. Observe G_l and condition (1) not satisfied: Our approach characterizes this scenario as optimal choice relative to benchmark predictions. It is possible, however, that if condition (2) is also not satisfied, the scenario reflects conservative choice.

The above conditions imply that our characterization strategy may *overestimate* the share of optimal choice (i, iv) and *underestimate* the share of conservative (iv) and aggressive choice (i). In this sense the original estimates offer an upper bound of decision errors relative to benchmark predictions.

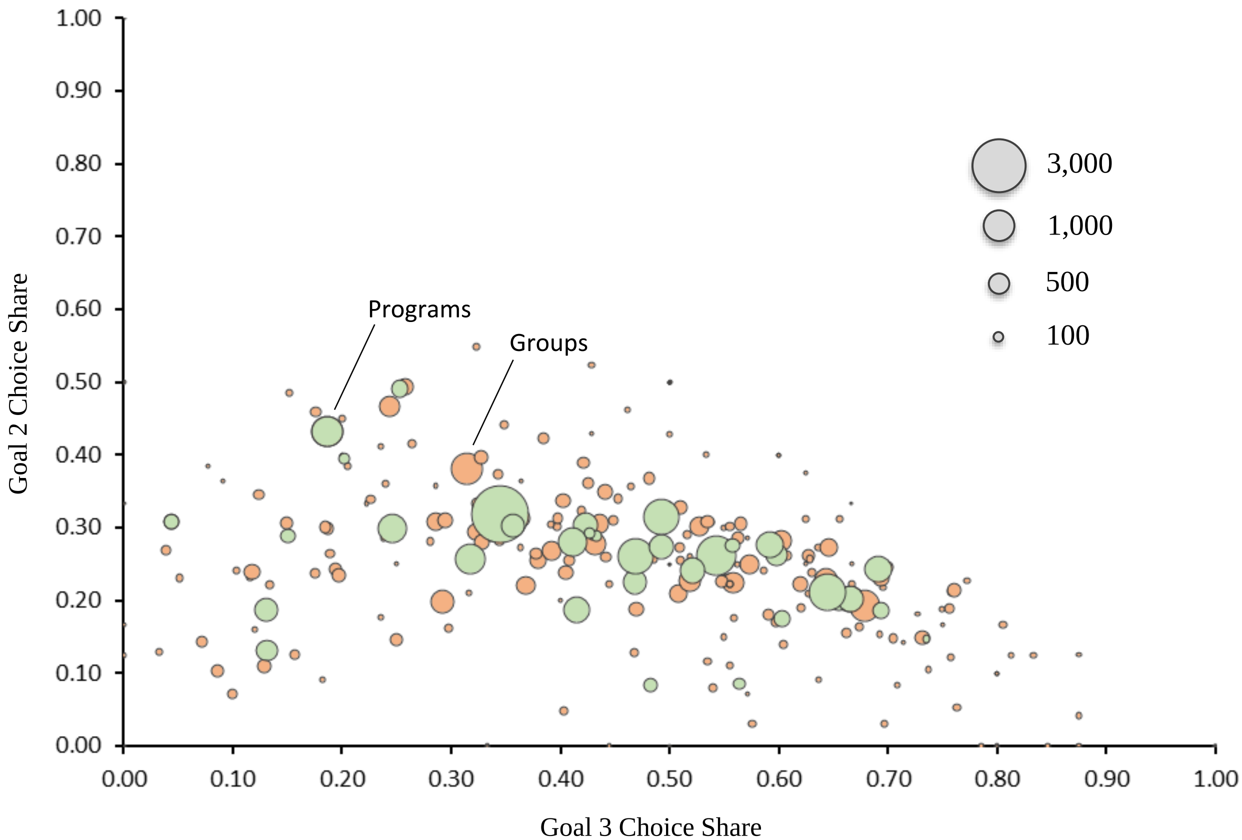
It is straightforward to modify this framework, which conceptualizes choice from a menu of two nested options, to accommodate a menu of three options such as that featured in GQ. For example, if one were to specify that employees first decide between Goals 1 and 2 and then decide between the superior of the two low goals and Goal 3, the modified framework would, once again, imply that the original estimates of optimal choice share should be interpreted as an upper bound while original estimates of the conservative choice share should be interpreted as a lower bound. In practice, because the empirical frequency of (ii) and (iv) is low in the data field data, we speculate that the original estimates of the conservative choice share are likely to be accurate.

APPENDIX – REFERENCES

Holt, C.A., & S.K. Laury. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92 (5): 1644-1655**.**

Post, T., Van den Assem, M. J., Baltussen, G., & R.H. Thaler. (2008). Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review*, 98(1): 38-71.

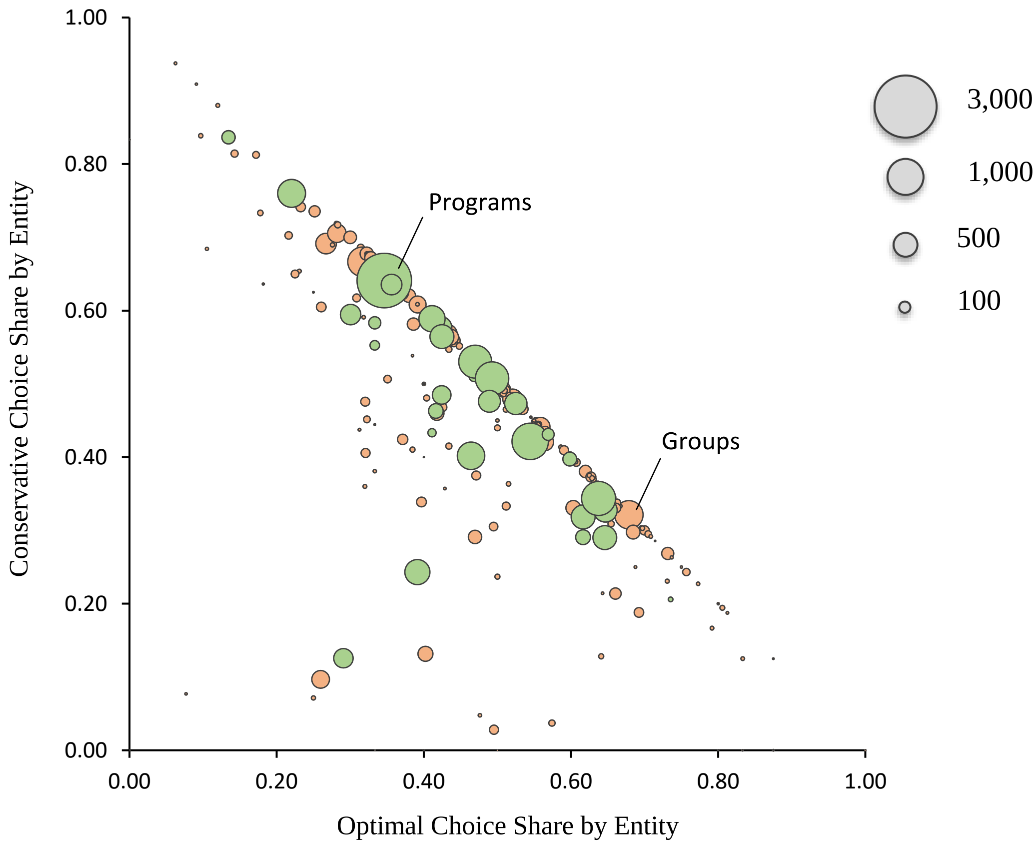
Figure 1.
Average Program and Group Choice Shares for Goals 2 and 3



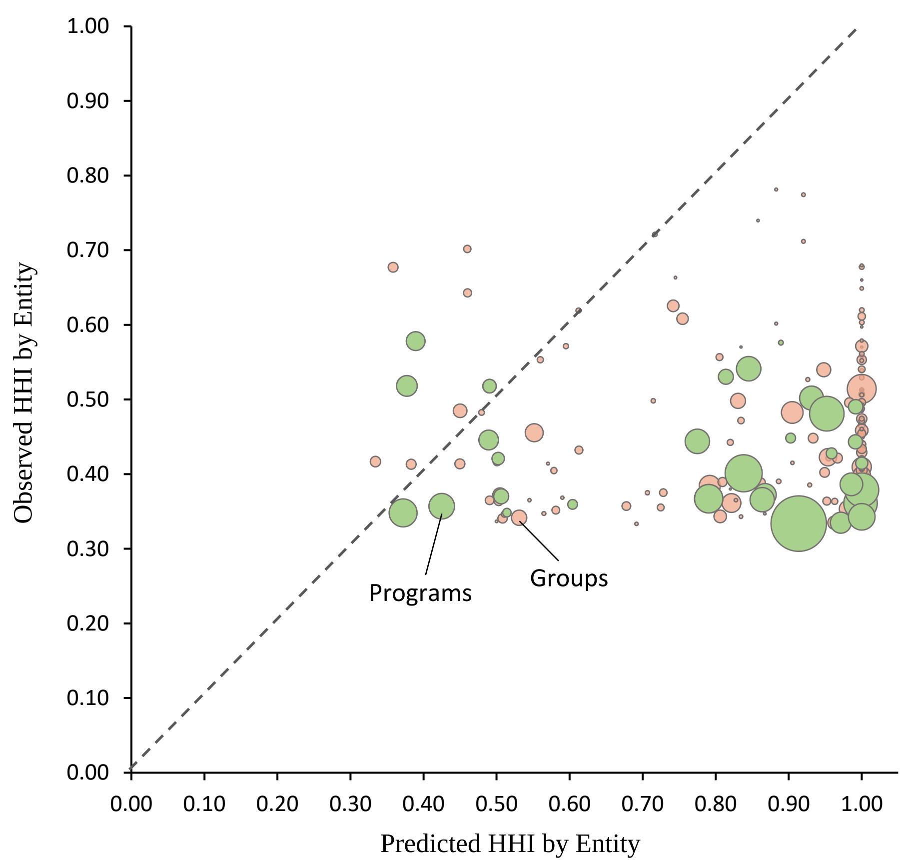
Notes: This figure depicts average choice shares for Goals 2 and 3 for each program (green) and group (orange). Groups with less than 10 employees excluded.

Figure 2.
Choice Characterization by Program and Group under Expected Value Benchmark

Panel A. Optimal and Conservative Choice Shares



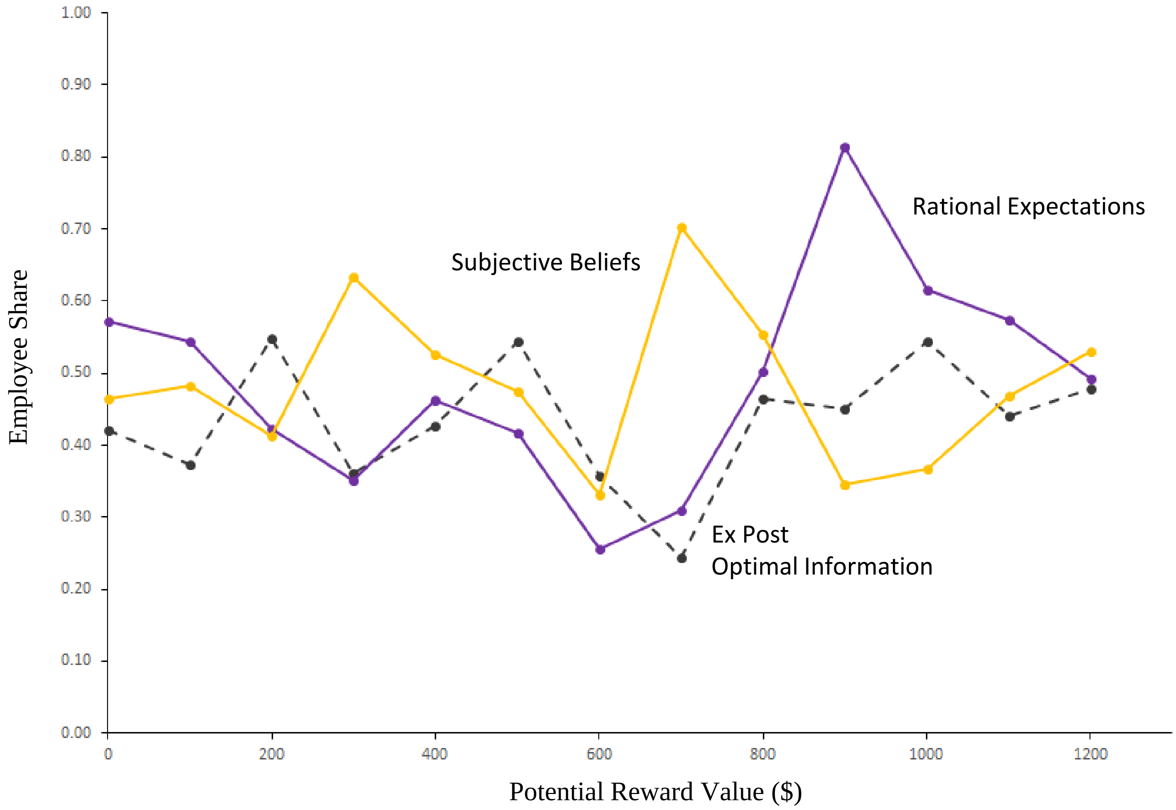
Panel B. Predicted and Observed Choice Heterogeneity (HHI)



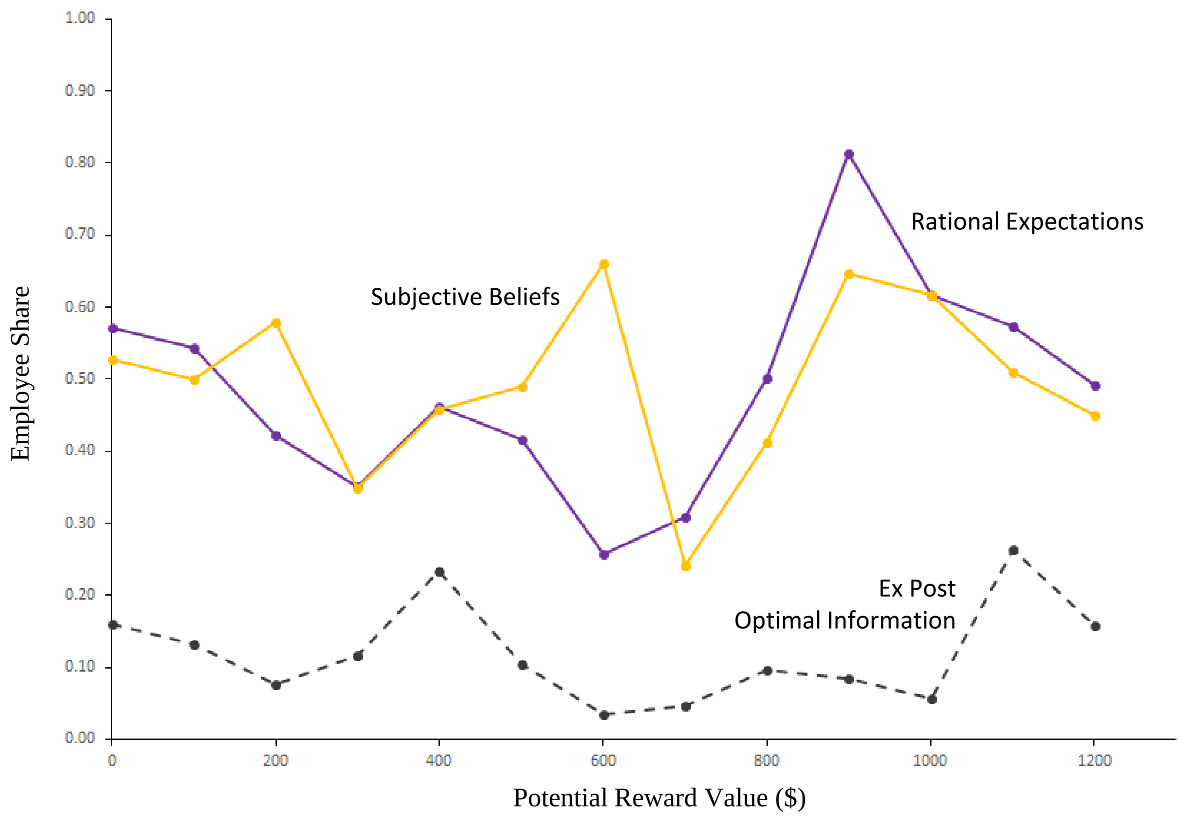
Notes: This figure depicts average optimal and conservative choice shares (Panel A) and predicted and observed choice heterogeneity as indicated by the Herfindahl-Hirschman Index (HHI) (Panel B) under an expected value benchmark for each program (green) and group (orange). Groups with less than 10 employees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

Figure 3.
Optimal and Conservative Choice under Expected Utility Benchmark by Potential Reward and Information Regime

Panel A. Optimal Choice Share under Expected Utility (r = 0.0003)



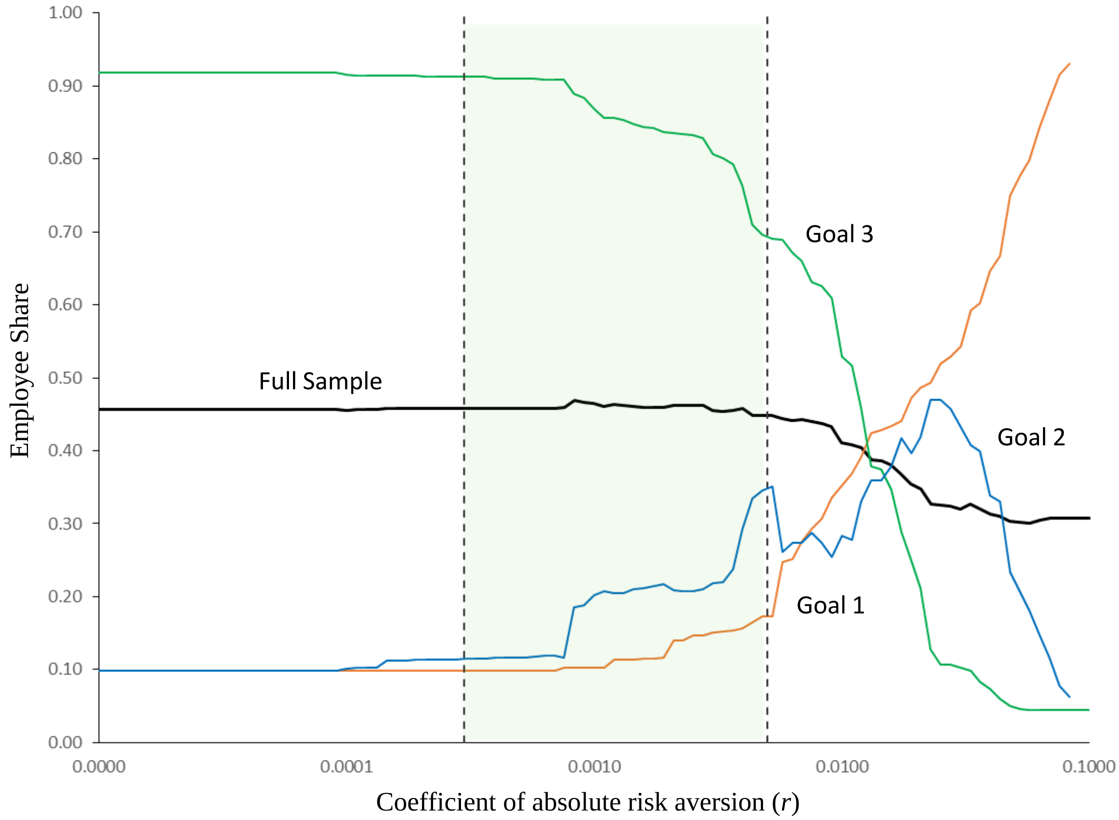
Panel B. Conservative Choice Share under Expected Utility (r = 0.0003)



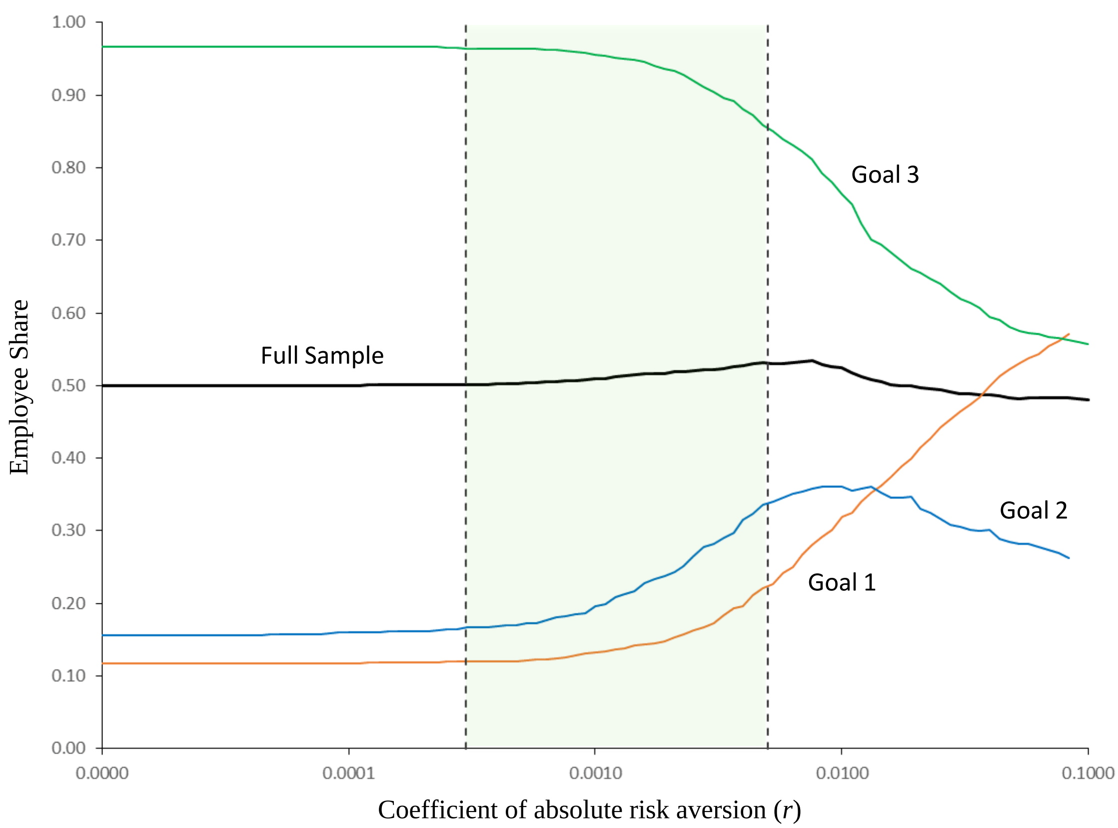
Notes: This figure reports the share of optimal (Panel A) and conservative (Panel B) choice by potential reward value under the expected utility (r = 0.0003) benchmark across varying assumptions regarding employee beliefs. Potential reward value refers to an employee's largest earnable reward (Goal 3) and is censored at $1,150.

Figure 4.
Optimal Choice under Expected Utility Benchmark by Risk Preference and Information Regime

Panel A. Expected Utility Benchmark assuming Rational Expectations

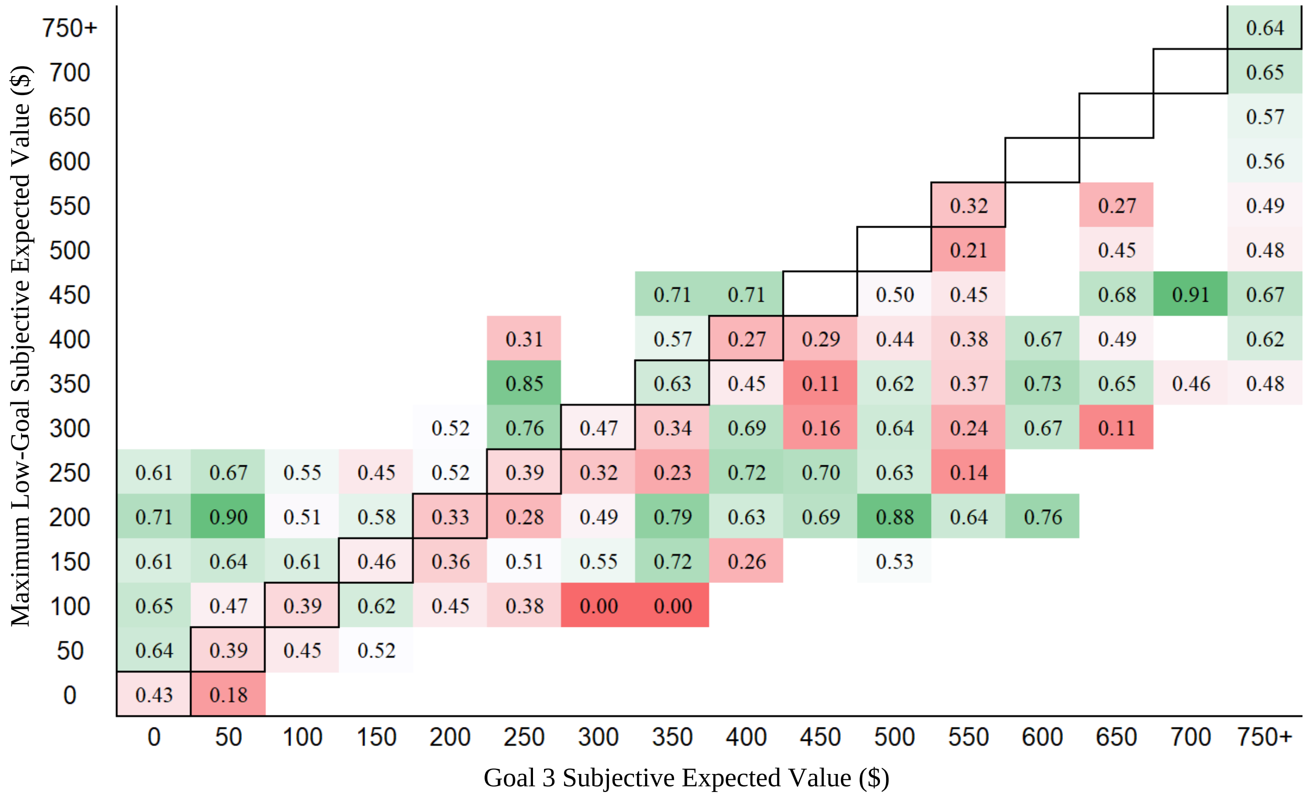


Panel B. Expected Utility Benchmark assuming Subjective Beliefs



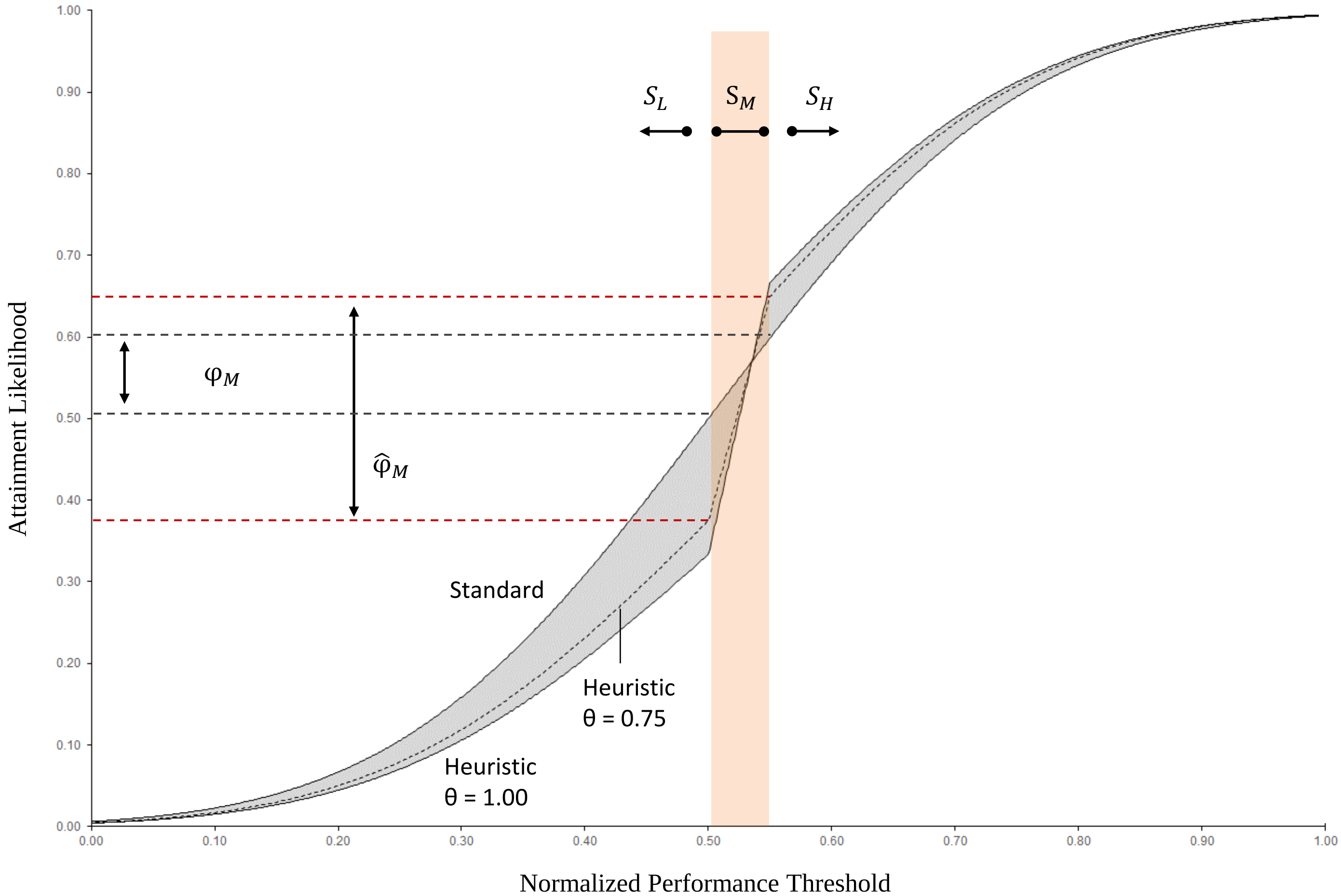
Notes: This figure depicts the share of optimal choice overall and by goal choice under expected utility across varying assumptions regarding CARA risk preferences and employee beliefs. Panel A depicts the share of optimal choice assuming rational expectations for an extended range of *r* on a logarithmic scale while Panel B depicts the analogous optimal choice share assuming subjective beliefs. The shaded region denotes the range of substantial but still plausible risk aversion, $r \in [0.0003, 0.005]$.

Figure 5.
Optimal Choice by Subjective Goal Value under EU Benchmark (r = 0.0003)



Notes: This figure depicts optimal choice shares under an expected utility benchmark across a matrix of subjective expected goal values. Subjective expected values are censored at $750 and cells with less than 10 observations are omitted. Low-goal subjective expected values refer to the maximum subjective expected value of Goals 1 and 2.

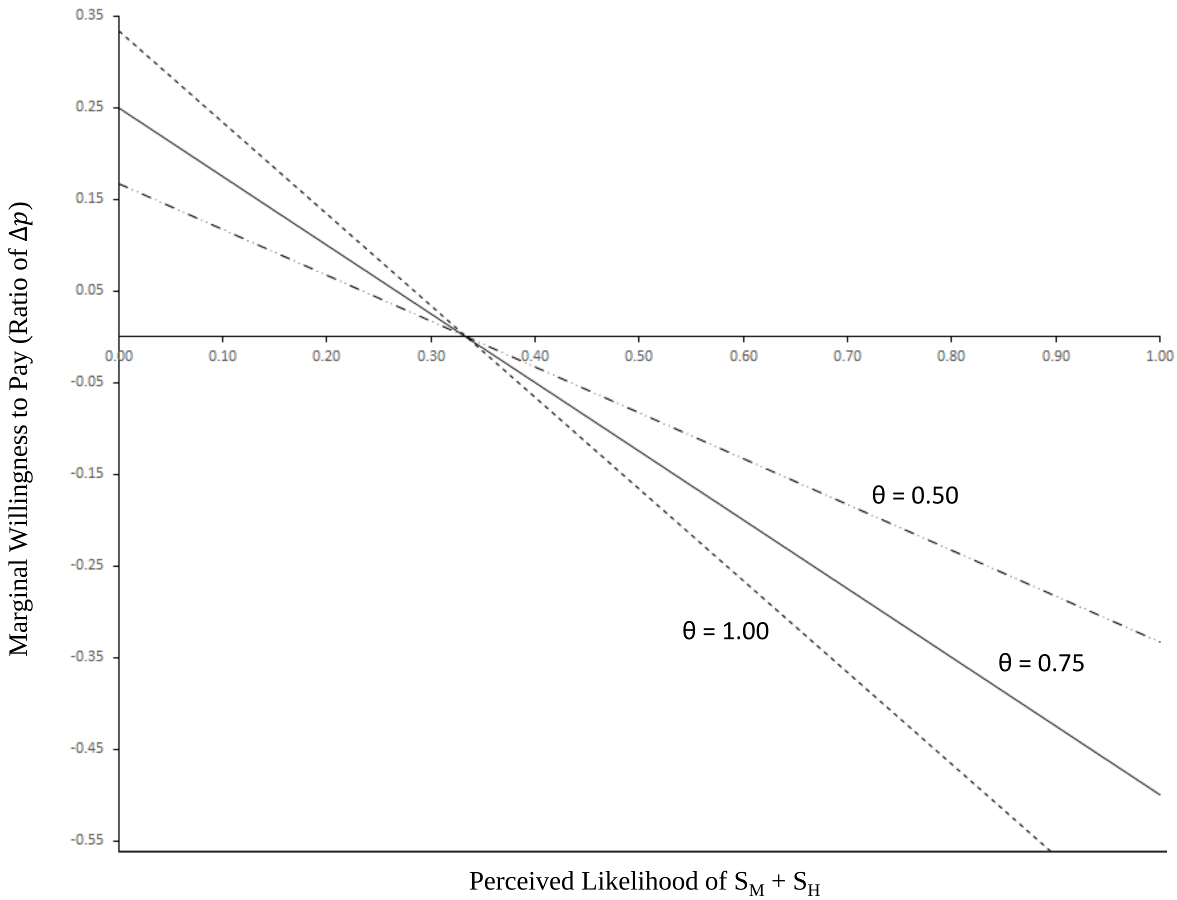
Figure 6.
Pairwise Partition Dependence and Perceived Goal Attainment CDFs



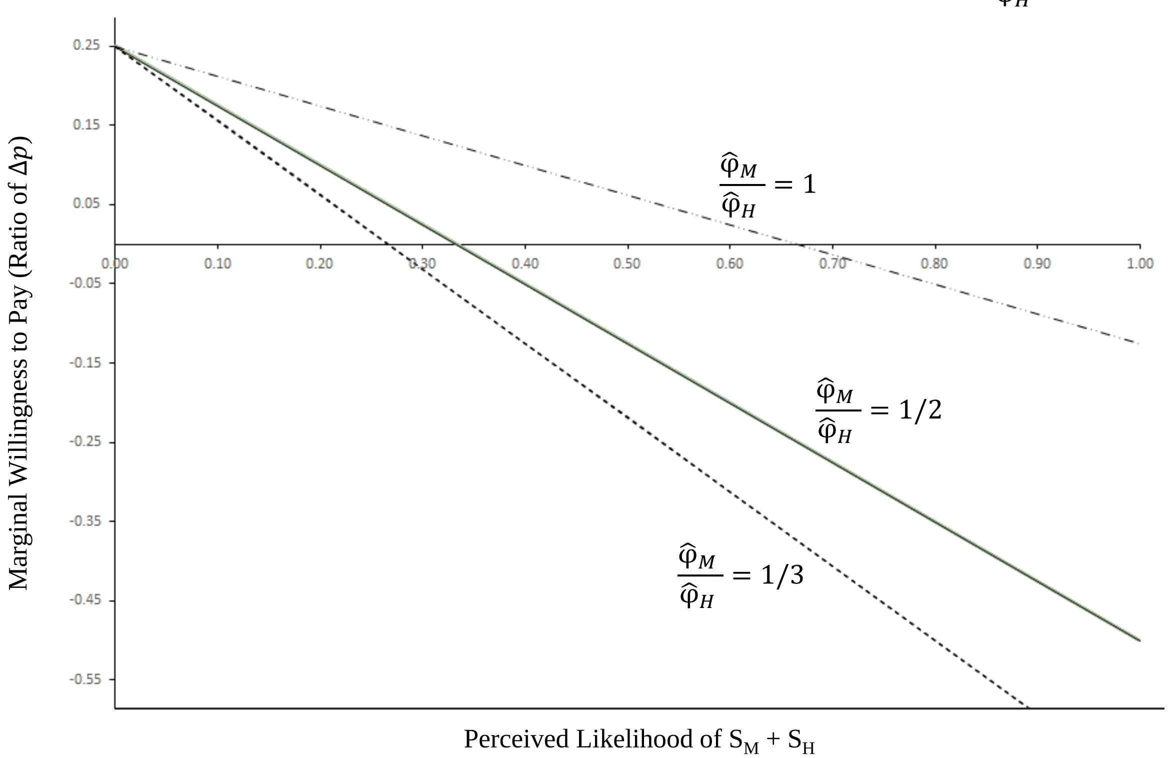
Notes: This figure depicts stylized goal attainment CDFs for GQ pairwise comparison under standard and heuristic choice (θ = [0.75, 1.00]). The shaded vertical region depicts the between-goal state space; the region to its left denotes the low-goal state space; and the region to its right denotes the high-goal state space. Shaded areas between CDFs depict inferential bias under heuristic evaluation. Arrows denote the actual and perceived likelihood of φ_M under standard and heuristic choice (θ = 0.75).

Figure 7.
Net Insurance Demand Bias under Heuristic Choice across Perceived Loss Likelihood

Panel A. Net Insurance Demand Bias and Bias Severity

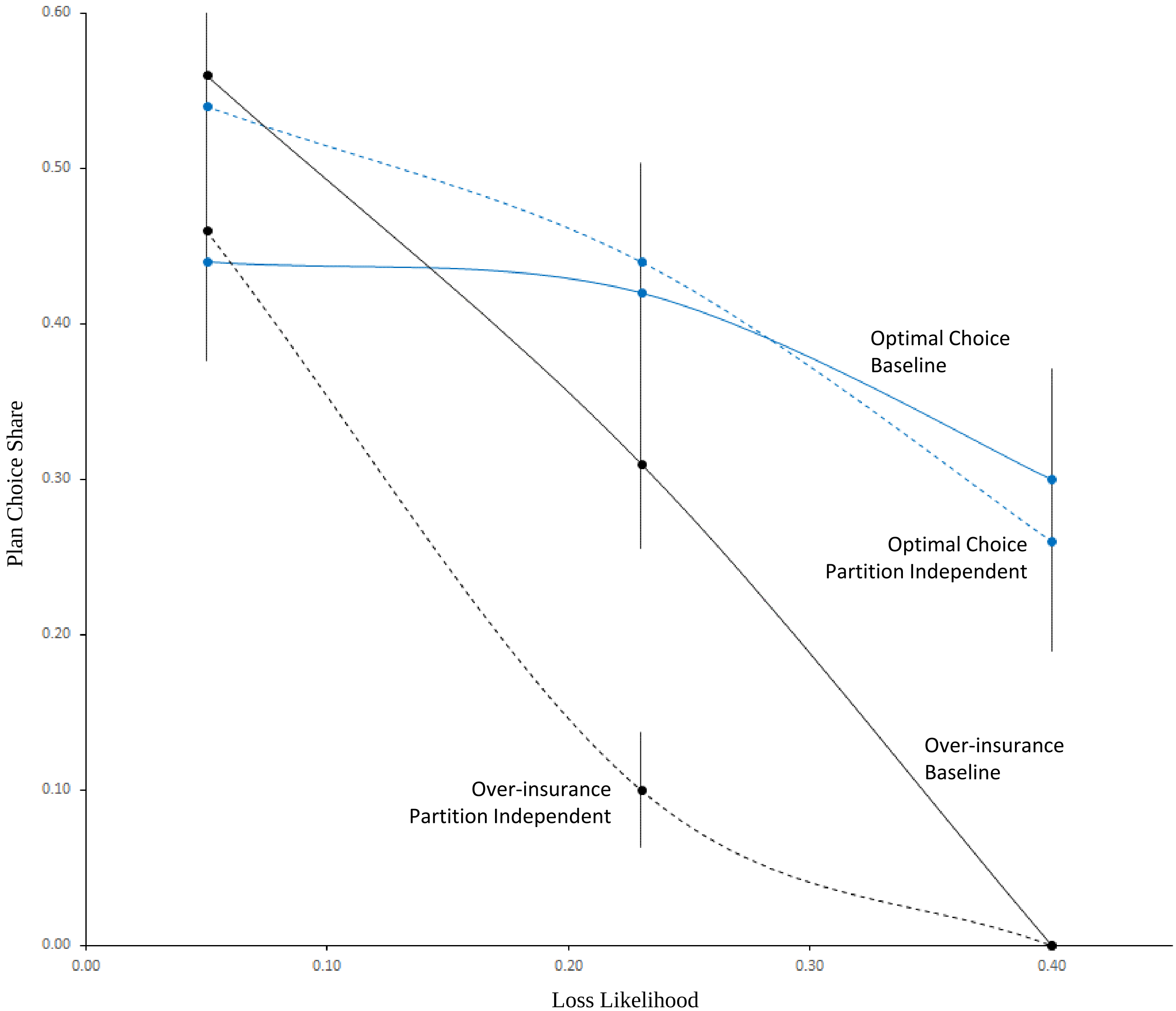


Panel B. Net Insurance Demand Bias and Relative Risk Ratio ($\frac{\widehat{\varphi}_M}{\widehat{\varphi}_H}$)



Notes: This figure depicts the net bias in insurance demand under heuristic choice across varying baseline levels of perceived risk loss for different bias severity (Panel B) and relative risk ratios (Panel B). Panel A assumes a relative risk ratio of 1:2 and Panel B assumes θ = 0.75. Net bias is expressed as the excess willingness to pay for a high versus low coverage plan, relative to a standard benchmark, as a ratio of the price difference between plans.

Figure 8.
Home Insurance Plan Choice across Loss Likelihood (Experiment F)



Notes: This figure depicts the share of over-insurance (black) and optimal choice (blue) for the baseline (solid) and partition independent (dashed) conditions across varying loss likelihood from Experiment F. Loss likelihood refers to the percent likelihood of any loss during the plan coverage period (i.e., 0.05, 0.23, or 0.40). Error bars denote +/- 1 standard error.

Table 1.
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 observable detail on GQ programs and employees for the primary sample. Panel A describes the number and size of programs, while Panel B describes employee-level statistics regarding average program duration and potential rewards. Potential reward value refers to an employee's largest earnable reward (Goal 3 reward). Panel C summarizes employee demographic detail overall and by sub-groups distinguished by potential reward value. We impute age from self-reported 10-year bins, infer gender using a combination of self-reported data and inferences from first name, and approximate salary with program-level averages for programs with available data.

Table 2.
Goal Choice, Employee Productivity, and Goal Attainment

	All	Sample Restricted by Goal Choice		
		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, and goal attainment for the primary sample overall and by employee goal choice. Panel A summarizes goal choice and average potential rewards, where potential reward value refers to an employee's largest earnable reward (Goal 3 reward). Panel B summarizes employee productivity both relative to baseline and to Goal 3 (the former measure excludes 18 percent of employees with no baseline data). Panel C summarizes goal attainment and average earned rewards.

Table 3.
Goal Choice Characterization under Expected Utility Benchmarks

	Risk Neutral EU		Expected Utility (CARA)					
			Rational Expectations			Subjective Beliefs		
	Rational	Subjective	r = 0.0003	r = 0.005	r [0, 0.005]	r= 0.0003	r = 0.005	r [0, 0.005]
Panel A. Characterization Overview								
Optimal Choice	0.45	0.50	0.45	0.44	0.56	0.50	0.53	0.59
Conservative Choice	0.49	0.48	0.49	0.38	--	0.48	0.39	--
Aggressive Choice	0.06	0.02	0.06	0.17	--	0.02	0.08	--
Predicted Heterogeneity [Herfindahl-Hirschman Index]	0.72	0.76	0.70	0.45	--	0.75	0.56	--
Gender Gap in Conservative Choice [F - M]	0.14	0.12	0.14	0.11	0.11	0.12	0.11	0.11
Panel B. Counterfactual Loss | Conservative Choice								
Realized Reward	164	162	162	122	--	159	118	--
Counterfactual Reward | Ex Ante Optimal Choice	303	281	302	244	--	276	222	--
Loss as % of Counterfactual Reward	0.46	0.42	0.46	0.50	--	0.42	0.47	--
Loss as % of Realized Reward	0.85	0.73	0.86	1.00	--	0.74	0.88	--
Panel C. Optimal Choice Share by Reward and Tenure								
Potential Reward Value								
Highest Quartile	0.42	0.48	0.42	0.39	0.67	0.49	0.55	0.72
Lowest Quartile	0.44	0.48	0.44	0.44	0.44	0.48	0.48	0.45
Employee Tenure								
Highest Category [10+ Years]	0.39	0.45	0.40	0.40	0.55	0.46	0.53	0.61
Lowest Category [< 1 Year]	0.44	0.47	0.44	0.44	0.52	0.47	0.50	0.54

Notes: This table characterizes the efficiency of goal choice for the primary sample under expected utility across varying assumptions regarding CARA risk preferences and employee beliefs. Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the specified benchmark models. It also reports the implied female-male gender gap in conservative choice and the degree of predicted choice heterogeneity associated with each benchmark. Panel B summarizes the economic consequence of conservative goal choice for employees attaining the lowest goal. Panel C reports the share of optimal choice across employee sub-groups distinguished by potential reward value and employee tenure. Blank cells reflect an inability to uniquely characterize aggressive and conservative choices for benchmarks involving flexible values of r.

Table 4.
Employee Beliefs and Confidence of Goal Attainment

	All	By Goal Choice		
		Goal 1	Goal 2	Goal 3
Panel A. Beliefs of Goal Attainment				
Rational Expectations				
Goal 1	0.44	0.41	0.44	0.46
Goal 2	0.37	0.32	0.36	0.39
Goal 3	0.30	0.25	0.28	0.33
Subjective Beliefs				
Goal 1	0.78	0.65	0.79	0.86
Goal 2	0.69	0.50	0.71	0.82
Goal 3	0.63	0.43	0.57	0.77
Panel B. Over/Under Confidence				
Ratio of Subjective/Rational Beliefs				
Goal 1	2.20	2.09	2.26	2.27
Goal 2	2.62	2.42	2.79	2.76
Goal 3	3.46	3.26	3.43	3.59
Relative Ratio of Over/Under Confidence				
Goal 3/Goal 1	1.45	1.41	1.42	1.48
Goal 3/Goal 2	1.22	1.24	1.15	1.22
Goal 2/Goal 1	1.13	1.08	1.18	1.17

Notes: This table summarizes employee beliefs and confidence associated with goal attainment for the primary sample overall and by employee goal choice. Panel A summarizes goal-attainment beliefs under rational expectations and subjective beliefs (see text for detail on estimates of rational expectations). Subjective beliefs reflect employee self-reports elicited during enhanced enrollment using a scale from 0 to 100 percent with 10-percent increments. For tractability, we adjust beliefs of 0 and 100 percent to 1 and 99 percent, respectively. Panel B summarizes average employee under/over confidence for each goal and relative under/over confidence for goal pairs. We represent confidence by the average ratio of subjective beliefs and rational expectations (> 1 indicates overconfidence), winsorized at the 5th and 95th percentiles.

Table 5.
Goal Choice Characterization under Non-Standard Benchmarks

	SEU Baseline (CARA, r = 0.0003)	Non-Linear Weights [Prelec, $\alpha = \beta = 0.65$]	Composite Gain-Loss [RP = g; $\eta = 1$; $\lambda = 2.25$]
<u>Panel A. Characterization Overview</u>			
Optimal Choice	0.50	0.47	0.59
Conservative Choice	0.48	0.52	0.24
Aggressive Choice	0.02	0.01	0.17
Predicted Heterogeneity [Herfindahl-Hirschman Index]	0.75	0.84	0.41
Gender Gap in Conservative Choice [F - M]	0.12	0.12	0.06
<u>Panel B. Counterfactual Loss | Conservative Choice</u>			
Realized Reward	159	163	168
Counterfactual Reward | Ex Ante Optimal Choice	276	282	296
Loss as % of Counterfactual Reward	0.42	0.42	0.43
Loss as % of Realized Reward	0.74	0.73	0.76
<u>Panel C. Optimal Choice Share by Reward and Tenure</u>			
Potential Reward Value			
Highest Quartile	0.49	0.44	0.61
Lowest Quartile	0.48	0.46	0.55
Employee Tenure			
Highest Category [10+ Years]	0.46	0.42	0.59
Lowest Category [< 1 Year]	0.47	0.44	0.59

Notes: This table characterizes the efficiency of goal choice for the primary sample under non-standard benchmark models. The first column provides a baseline characterization of choice under the SEU benchmark (CARA utility, r = 0.0003); the second column characterizes choice for a modified benchmark allowing for non-linear probability weights (Prelec 1998); the final column characterizes choice for the best-performing gain-loss utility benchmark (see text for details). Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the specified benchmark models. It also reports the implied female-male gender gap in conservative choice and the degree of predicted choice heterogeneity associated with each benchmark. Panel B summarizes the economic consequence of conservative goal choice for employees attaining the lowest goal. Panel C reports the share of optimal choice across employee sub-groups distinguished by potential reward value and employee tenure.

Table 6.
Goal Choice Characterization under Expected Utility and Heuristic Benchmarks — Experiment A

	SEU Baseline (CARA, r = 0.0003)	Non-Linear Weights [Prelec, α = β = 0.5]	Composite Gain-Loss [RP = g; η = 1] [λ = 2.25]	Personal λ	Flexible λ	Contextual Sorting Heuristics: Ability	Contextual Sorting Heuristics: Taste for Competition
All Menus (6/6)	0.04	0.03	0.18	0.19	0.29	0.10	0.09
Nearly All Menus (5+/6)	0.16	0.13	0.31	0.30	0.44	0.12	0.10
All 3 Goal Menus (4/4)	0.16	0.09	0.25	--	--	0.13	0.12
All 4 Goal Menus (2/2)	0.10	0.10	0.32	--	--	0.15	0.15

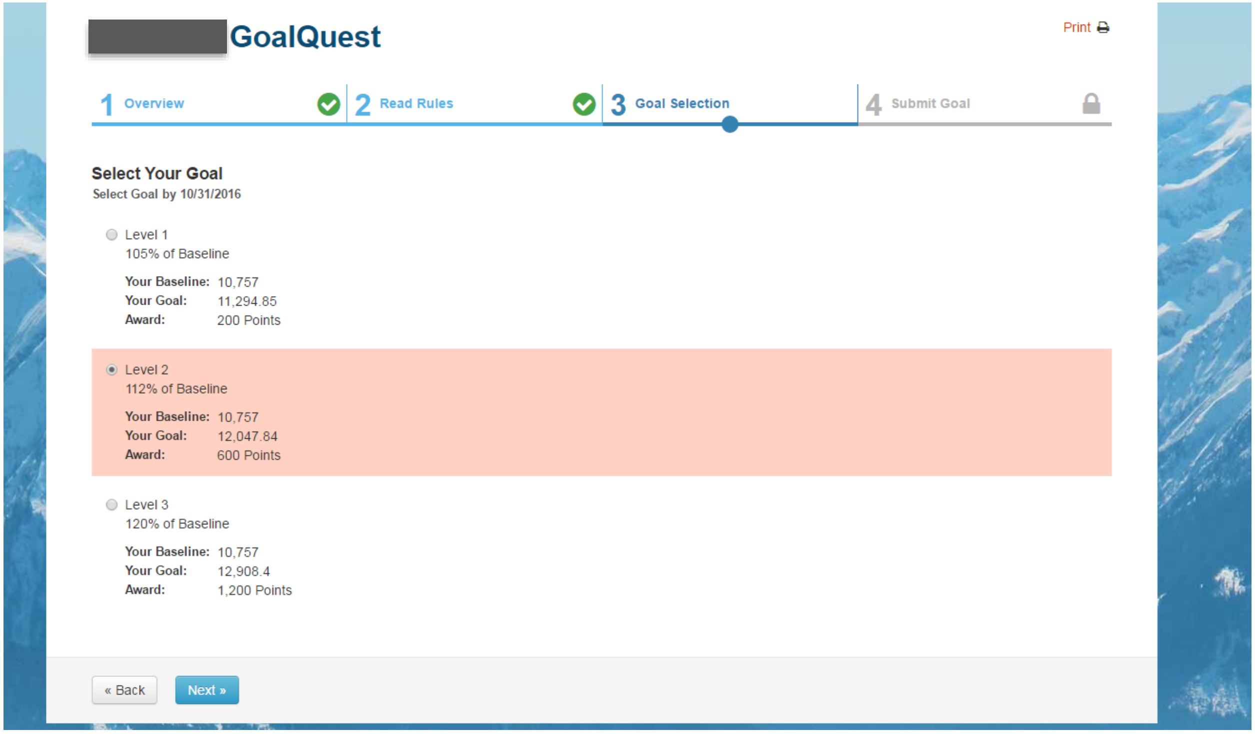
Notes: This table characterizes multiple measures of optimal goal choice for experimental participants under a range of benchmark models (Experiment A). The first column provides a baseline characterization of choice under the SEU benchmark (r = 0.0003), while the next few columns characterize choice for benchmarks modified to allow for non-linear decision weights and gain-loss utility. A final set of columns characterizes choice for heuristic models involving contextual sorting by self-reported ability or tastes for competition.

Table 7.
Goal Choice Characterization under PDD Heuristic in the Lab and Field

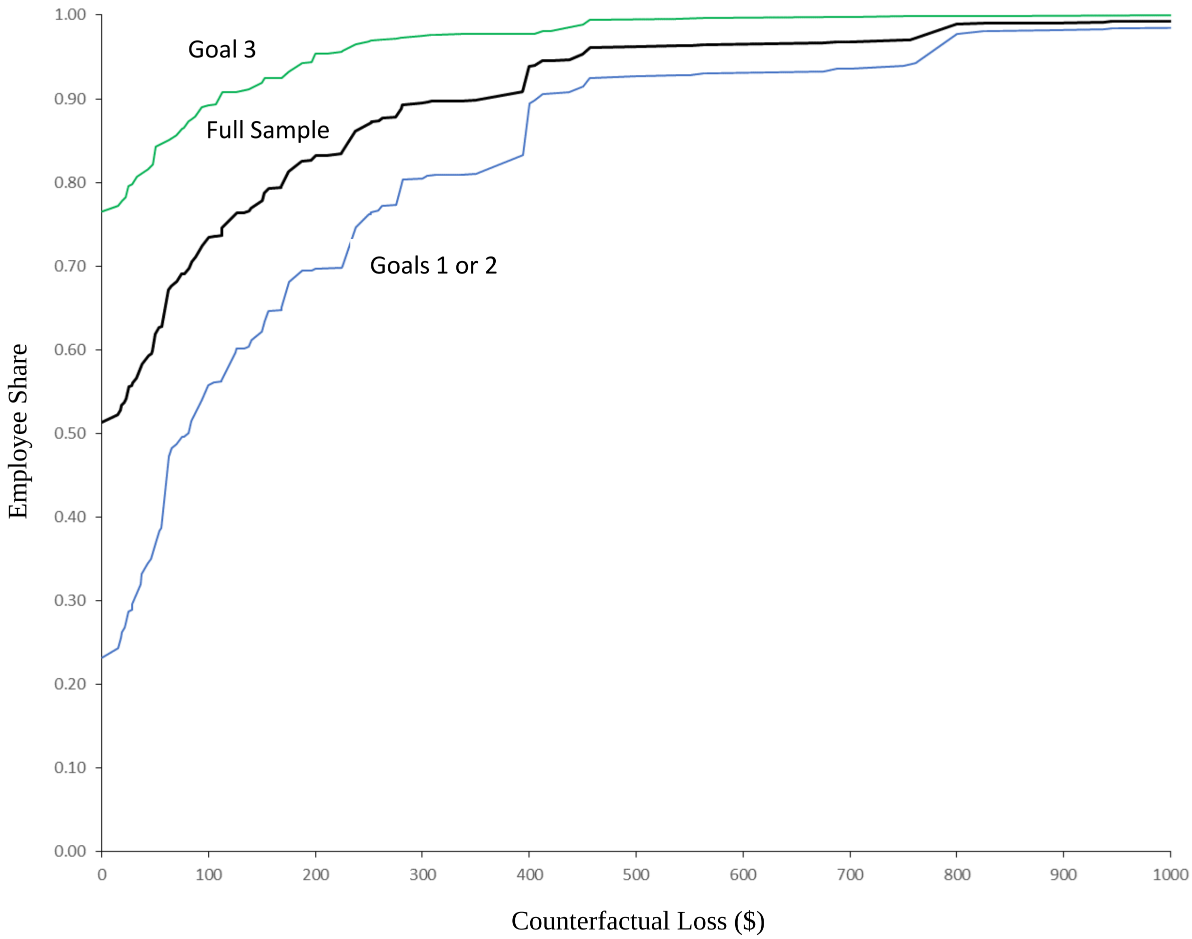
		PDD Heuristic - Model Parameters								
		No Bias			Personalized Bias			Parameterized Bias (θ = 0.75)		
Decision Sample	RN SEU	[w = $0]	[w = $25; 1¢]	[w = $50; 2¢]	[w = $0]	[w = $25]	[w = $50]	[w = $0]	[w = $25; 1¢]	[w = $50; 2¢]
Field Data	0.50	0.51	0.65	0.78	--	--	--	0.47	0.83	0.92
Experiment C	0.38	0.38	0.42	0.45	0.56	0.59	0.72	0.36	0.49	0.63
Experiment A										
All Menus (6/6)	0.04	0.14	0.27	0.33	--	--	--	0.11	0.38	0.59
Nearly All Menus (5/6)	0.16	0.29	0.45	0.56	--	--	--	0.30	0.71	0.86

Notes: This table characterizes the optimality of goal choice in the primary field sample and lab (Experiments C and A) under a baseline benchmark and various formulations of the PPD heuristic benchmark. The first column provides a baseline characterization of choice under a risk-neutral subjective EU benchmark. The subsequent columns characterize heuristic choice across varying specifications of bias and noise. In columns in which two noise allowances are listed, the first pertains to the allowance in the field and Experiment C, while the second pertains to Experiment A.

Appendix Figure A1.
Sample Image of GQ Goal Selection Interface

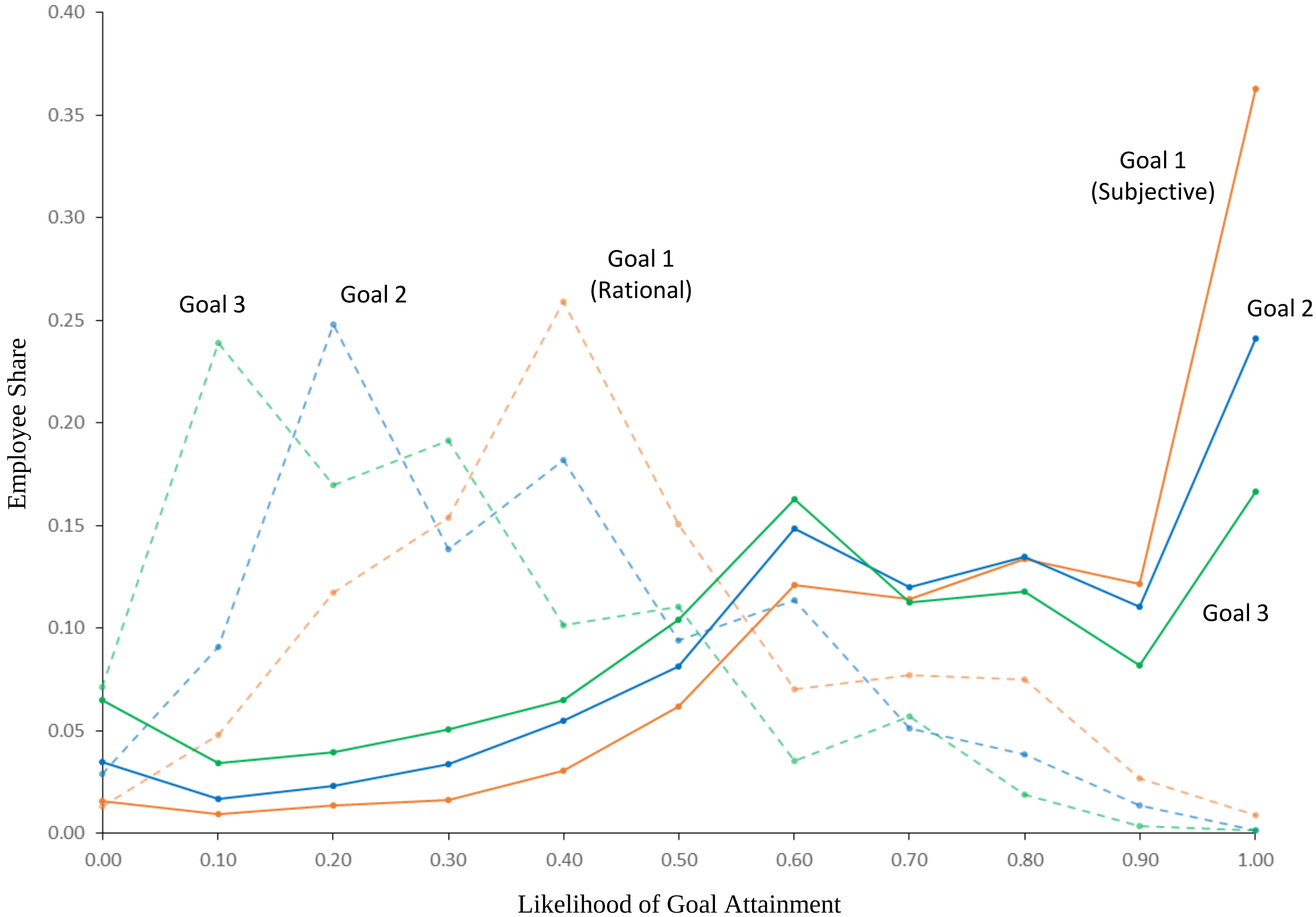


Appendix Figure A2.
Cumulative Distribution of Counterfactual Loss relative to Ex Post Optimal Choice | Goal Attainment



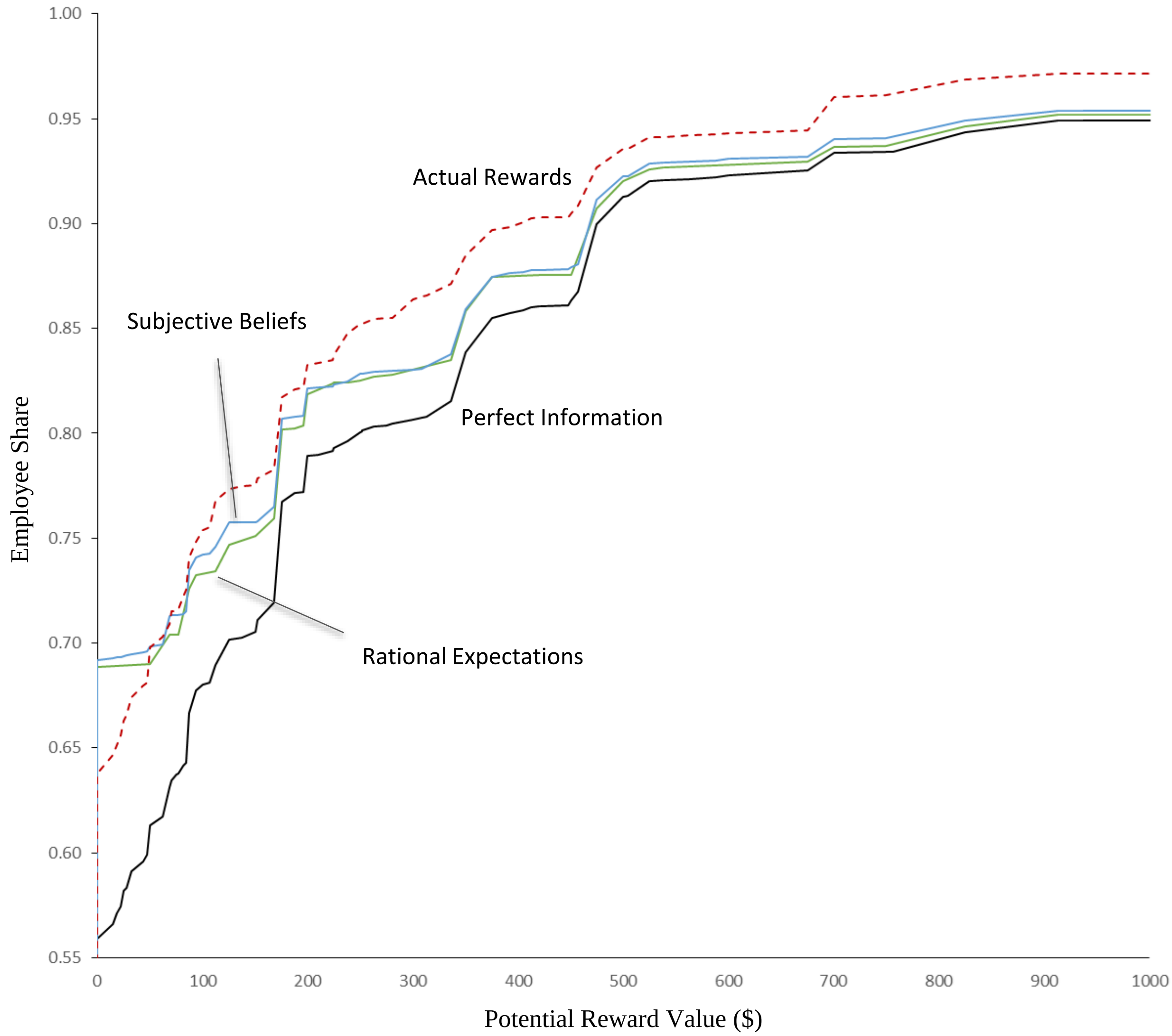
Notes: This figure depicts the cumulative distribution of counterfactual losses overall and by goal choice for employees whose productivity met or exceeded the Goal 1 threshold. Counterfactual loss refers to the difference between an employee's realized reward and the counterfactual reward an employee would have earned with ex post optimal choice (assuming no change in productivity). While counterfactual loss is censored at $1,000, a few employees had losses above $1,000 with a maximum loss of approximately $2,800.

Appendix Figure A3.
Distribution of Rational Expectations and Subjective Beliefs of Goal Attainment



Notes: This figure compares the distributions of rational expectations and subjective beliefs of goal attainment for each goal. We assign employee- and goal-specific rational expectations by adjusting the ex post average rate of goal attainment at the group-level by employee age and gender, as estimated from a linear regression (for a small share of employees, for whom this strategy violated monotonicity, we adopted the unadjusted ex post average). Subjective beliefs for each goal reflect employee self-reports, elicited during enhanced enrollment, using an eleven-point scale (0, 10, 20,…, 100 percent). For ease of comparison, the figure groups rational expectations into bins that parallel the subjective belief data.

Appendix Figure A4.
Cumulative Distribution of Actual and Counterfactual Rewards under Risk Neutral Expected Utility Benchmark by Information Regime



Notes: This figure depicts the cumulative distribution of actual rewards and counterfactual rewards under expected utility for risk neutral employees by information regime. Specifically, the dashed red line indicates the distribution of actual earned rewards, the black line indicates the distribution of counterfactual rewards given ex post optimal choice, the green line indicates the distribution of counterfactual rewards given ex ante optimal choice assuming rational expectations, while the blue line indicates the distribution of counterfactual rewards given ex ante optimal choice assuming subjective beliefs. The figure truncates the y axis at 0.55 to reflect the significant share of employees that did not attain any goal and, for clarity, truncates the x-axis at $1,000.

Appendix Table A1.
Goal Choice Characterization under Expected CRRA Utiity 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 characterizes the efficiency of goal choice with respect to EU benchmark models with CRRA utility across varying initial lifetime wealth and relative risk aversion. The second column reports the certainty coefficient (i.e., certainty equivalence as a share of expected value) assuming initial wealth of $25,000 for a fair bet of ($0, $10k). Highlighted region denotes interval of plausible relative risk aversion as indicated by Holt and Laury (2002). The first panel characterizes choice assuming rational expectations while the second panel characterizes choice assuming subjective expectations.

Appendix Table A2.
Optimal Goal Choice Shares under RN-SEU 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 optimal goal choice under a risk-neutral subjective EU benchmark model assuming varying specifications of effort costs. Baseline effort cost increment refers to the increase in hourly effort cost for Goal 2 versus Goal 1 as a % of wage. The convexity parameter refers to the proportional increase in effort costs for Goal 3 versus Goal 2 relative to the baseline increment. All calculations assume wage of $25/hour, 8 working hours per day, and the subjective beliefs elicited from employees. For example, a one-month program (~25 working days), baseline increment of 10%, and k = 1.5, implies total effort costs of $0, $500, and $1,250 for Goals 1, 2, 3, respectively.

Appendix Table A3.

Goal Choice Characterization for Expected Utility Benchmarks - Primary versus Expanded Sample

		Expected Utility (CARA)		
Characterization Overview	Risk Neutral	r = 0.0003	r = 0.005	r [0, 0.005]
Panel A. Primary Sample				
Optimal Choice	0.45	0.45	0.44	0.56
Conservative Choice	0.49	0.49	0.38	--
Aggressive Choice	0.06	0.06	0.17	--
Panel B. Expanded Sample				
Optimal Choice	0.41	0.41	0.40	0.54
Conservative Choice	0.56	0.55	0.48	--
Aggressive Choice	0.04	0.04	0.11	--

Notes: This table characterizes the efficiency of goal choice for the primary (Panel A) and expansive (Panel B) samples under expected utility across a range of assumptions regarding CARA risk preferences and assuming rational expectations.

Appendix Table A4.
Optimal Goal Choice Shares for Gain-Loss Utility Benchmarks by Candidate Reference Point

	Gain-Loss Utility (α = 0.88; η = 0)			Consumpton + Gain-Loss Utility (λ = 2.25)		
Candidate Reference Points	λ = 1.50	λ = 2.25	λ = 3.00	η = 1	η = 3	η = 5
<u>Panel A. Prospect Independent</u>						
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
<u>Panel B. Prospect-Dependent</u>						
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 assesses the descriptive accuracy of benchmark models involving gain-loss utility across several candidate reference points, functional forms, and parameter specifications. The first set of columns characterizes choice under benchmark models involving gain-loss utility following Kahneman and Tversky (1979) across potential values of the loss aversion parameter, λ. The second set of columns characterizes choice under benchmark models involving composite utility, an additively linear combination of consumption utility and gain-loss utility, across potential consumption utility scaling factors, n. (n = 0 therefore implies a model with gain-loss utility only). All benchmark models assume subjective beliefs. Panel A reports the share of optimal choice for prospect-independent candidate reference points while Panel B reports the analogous share of optimal choice for prospect-dependent candidate reference points. Please see text for additional detail on each of the benchmark models.

Appendix Table A5.

Demand for Prescription Drug Plans across Information Frames - Experiment D

	Menu Display		
	Baseline	Partition Dependent	Partition Independent
No Plan	0.11	0.18	0.13
Silver Plan [Coinsurance: 50%, Premium: $640]	0.59	0.53	0.44
Gold Plan [Coinsurance: 15%, Premium: $1220]	0.31	0.29	0.43
Expected Total Cost [Out-of-Pocket + Premium]	2076	2151	2041

Notes: This table reports average choice shares across conditions from Experiment D (N = 432). Participants were informed that coinsurance applies to all drug bills until the plan's out-of-pocket maximum of $7,500 (neither plan offered a deductible). They were also informed that annual drug bills could not exceed $10,000, even for those selecting no plan. Expected total cost refers to the estimated average total cost (premium + out-of-pocket costs) for participants in each condition based on their plan choices. Total cost estimates rely on an inferred distribution of potential drug bills (see text for details).

Appendix Table A6.
Demand for Home Insurance across Information Frames - Experiment E

	Menu Display			
	Full Information Baseline	No Information Baseline	Partition Dependent	Partition Independent
Basic Plan [Deductible: $1000, Premium: $616]	0.35	0.40	0.23	0.54
Medium Plan [Deductible: $500, Premium: $716]	0.39	0.41	0.50	0.38
Premium Plan [Deductible: $250, Premium: $803]	0.26	0.19	0.26	0.08
Expected Total Cost [Out-of-Pocket + Premium]	726	717	729	696

Notes: This table reports average plan choice shares across conditions from Experiment E (N = 435). Participants were informed that plans cover all expenses after the deductible has been met. Expected total cost refers to the estimated average total cost (premium + out-of-pocket costs) for participants in each condition based on their plan choice. Total cost estimates assume a 3 percent chance of damages exceeding $2500 and a 1 percent chance of damages of $500.