Risk Aversion in the Field:

Evidence on Prevalence and Motives from an Employee Rewards Program

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

Timothy Hyde
University of Pennsylvania
thyde@andrew.cmu.edu

This Version: October 2022

Abstract

Given its importance for theory, welfare, and policy, economists have long sought to understand the prevalence of and motives for risk aversion in the field. In practice, this inquiry is often confounded by the potential for biased beliefs (e.g., betting, investing), imperfect understanding (e.g., insurance), or limited generalizability (e.g., game shows). We overcome these challenges with new data detailing the choices, productivity, and beliefs of 20,133 employees across 18 large North American firms who participated in a simple, all-or-nothing, goal-rewards program with $9.4 million in incentives. First, we estimate that, under rational expectations, nearly one-half of employees select a goal lower than the EV-maximizing benchmark, resulting in an average unrealized reward of $272. Conservative goal choice persists across diverse financial stakes ($69 to $4,500) and employee experience. Second, we show that conservative choice cannot be explained by a standard expected utility (EUT) model with plausible levels of risk aversion or through commonly-discussed departures from EUT such as biased-beliefs (employees exhibit substantial *overconfidence* about productivity), non-linear decision weights, or gain-loss utility. We replicate the pattern of conservative choice and corroborate limits of EUT-based explanations through an incentive-compatible online goal-reward paradigm. We conclude by proposing, and experimentally testing, a novel decision-heuristic in which conservative choice emerges from the neglect of contingent probabilities in the context of proximal pairwise comparisons. We speculate that heuristic contingency neglect offers a potential explanation for the present evidence and for empirical puzzles in other menu-based settings such as those involving investing or insurance.

Acknowledgements: We extend a special thanks to Tim Houlihan for his insights and support and for facilitating access to these data. We additionally thank Linda Babcock, Karna Basu, Lynn Conell-Price, Stefano DellaVigna, Russell Golman, Kareem Haggag, Alex Imas, Botond Koszegi, George Loewenstein, Ted O'Donoghue, Lowell Taylor, Silvia Saccardo, Justin Sydnor, Oleg Urminsky and seminar participants at 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 Jenn Kelby, Mark Hirschfeld, and Betsy Schneider. Finally, we thank Stephanie Rifai, Cassandra Taylor, and a team of research assistants for excellent project support. The authors had full editorial discretion and any errors should be attributed 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, estimates of consumer welfare, and the optimal design of programs and policy. For example, in theory, assumptions about risk preferences should inform how policymakers regulate asset and insurance markets, how firms design contingent employee contracts, how individuals allocate investments, and how economists evaluate the welfare consequences of policies and programs. From the perspective of classical economic theory, risk-aversion among fully-informed, utility-maximizing individuals should reflect the diminishing marginal utility of wealth generated by the concavity of one's utility function (von Neumann and Morgenstern 1947).

The empirical evidence on financial decisions under risk (and possibly uncertainty), however, gives rise to a number of ostensible challenges for the standard model.[1] As one notable example, people exhibit a degree of risk aversion across a range of economic outcomes that is difficult to explain through the expected-utility framework. In the lab, researchers have documented risk aversion in the context of small-to-medium sized gambles implying an implausible degree of risk aversion at larger scales (Rabin 2000). And in the field, researchers have catalogued several examples of seemingly excessive risk-aversion in menu-based settings involving portfolio allocation or (home, auto, casualty, rental) insurance choice.[2] As a second example, many have noted that the observed variation in risk perceptions, or even preferences for risk, struggle to explain the observed heterogeneity in risky choice.

Over the last few decades, researchers have advanced a number of departures from the standard model of expected utility to explain risk aversion through alternative channels such as biased beliefs, non-linear decision weights (e.g., Kahneman and Tversky 1979; Prelec 1998), and loss (or disappointment) aversion in the context of gain-loss utility (e.g., Kahneman and Tversky 1979; Koszegi and Rabin 2006; Gul 1991; Loomes and Sugden 1986). As illustration, risk averse financial choice could reflect an overestimation of underlying risk, disproportionate weighting of unlikely outcomes, or an elevated willingness to pay to avoid unexpected out-of-pocket costs. Risk aversion could also emerge from non-standard processes receiving less attention in economics such as those involving heuristics, salience, affect, cognitive processes (e.g., fuzzy-trace theory), or hormones (see, e.g., Kusev et al., 2017; Fox, Erner, and Walters, 2015). Practically, clarifying motives for risk-taking in the field has been complicated

[1] For simplicity, we largely elide the important distinction between risk (knowledge of the probability distribution over potential outcomes) and uncertainty (a lack of knowledge of the probability distribution over potential outcomes) in the paper. We briefly engage the implications of uncertainty when we discuss allowances for error in beliefs.

[2] Conditioned on purchasing any policy, research has documented several examples of implied risk aversion difficult to rationalize through standard expected utility (e.g., Cutler and Zeckhauser, 2004; Sydnor 2010; Kunreuther et al. 2013).

by limited data on beliefs (e.g., betting markets, insurance), complexity of the choice environment (e.g., insurance, financial markets), or limited generalizability (e.g., game shows).

We overcome several of these challenges by analyzing an unusually rich dataset describing the decisions—and beliefs—of employees in the context of a simple, all-or-nothing, employee goal-rewards program. The program, known as GoalQuest© (GQ), was conceived by BI Worldwide (BIW), a US-based consulting firm specializing in the design and administration of programs that leverage principles of behavioral science to increase employee/consumer engagement. At the onset of each program, typically lasting one to three months, participating employees are asked to privately self-select a productivity goal from a menu of three options, personalized based on their productivity during a pre-program control period. Critically, each goal corresponds to an often substantial all-or-nothing reward (i.e., selecting Goal 3 but achieving Goal 2 would result in no reward) denominated in points redeemable for a non-monetary prize according to a preset exchange rate. To encourage high goal choice, most goals increased linearly (e.g., Goal 1: 100 units, Goal 2: 110 units, Goal 3: 120 units) while the value of rewards almost always increased non-linearly (e.g., Goal 1: $100, Goal 2: $300, Goal 3: $600).

Our primary evidence on financial risk aversion draws from the goal choices and beliefs regarding goal attainment of 20,133 employees who participated in 34 distinct GQ programs administered across 18 large North American (primarily US) firms from 2016 to 2019. These employees stood to earn $9.4 million in potential rewards through the program, or an average of $467. We corroborate our results with additional data describing the goal choices and productivity, but not beliefs, of another 18,528 employees who stood to earn another $8.5 million. To organize potential explanations for conservative goal choice, we outline a simple framework in which a fully-informed, utility-maximizing, risk-neutral, employee must select either a high or low goal associated with an all-or-nothing reward. We then successively introduce departures from this baseline—preference-based risk aversion, biased beliefs, non-linear decision weights, and gain-loss utility—to produce a series of benchmark models that we assess with the data. To explore additional mechanisms and potential confounds, we supplement our analysis from the field with evidence an incentive-compatible goal-rewards paradigm in the context of an online effort-task. We conclude by proposing, and testing, a novel heuristic choice strategy to explain financial risk-taking in this setting and perhaps more broadly.

For several reasons, we see this setting as an attractive litmus test for understanding how people engage financial risk. First, the findings should be highly generalizable given the diversity of the sample, the near-complete participation rate, and the wide-ranging financial stakes. Specifically, these data describe the behavior of employees across a diversity of age, gender, occupation, industry, geography, experience, and salary who participate in the program at a rate in excess of 98 percent, suggesting a high degree of ecological validity. And while our analysis is restricted to large programs administered during a

few-year period, these programs share the same structure as other GQ programs which have collectively been administered, since 2001, to a substantial share of Fortune 500 firms. The generalizability of the setting is also conveyed by the variance in reward value, from an estimated $69 to $4,500—an interval that encompasses many household financial decisions of interest to economists. Second, despite the aforementioned diversity across employees and stakes, GQ asks employees to make decisions from a highly standardized and simple choice menu. The simplicity of the decision environment stands in contrast to other more complex settings where attitudes towards risk may be difficult to infer. Finally, our partnership with BIW led to the development of an enhanced enrollment process through which we elicited contemporaneous employee beliefs of goal attainment. The data on beliefs were instrumental not only for testing mechanisms but also for ruling out potential confounds such as the potential differences in the cost of effort across employees.

Our analyses of employee decisions yield three main findings. A first finding is that employee goal choice exhibits a substantial degree of risk aversion and choice heterogeneity (Goal 1: 0.29, Goal 2: 0.27, Goal 3: 0.44). Assuming rational expectations, estimated using a procedure borrowed from the literature on insurance, implies 49 percent of employees selected a lower goal than that predicted by the expected-value (EV) maximizing benchmark (for most employees, Goal 3). Conservative goal choice, for employees who attained the low-goal threshold, resulted in an average unrealized reward of $139 (30 percent of the average potential reward) in rational expectation and $272 (58 percent of the average potential reward) based on an employee's ex post productivity. Overall, 45 percent of employees chose the EV-optimal goal, a share that does not meaningfully vary across the reward size, employee tenure, or approximate salary (implying conservative choice is not restricted to settings with financial illiquidity).

Our second finding is that plausible utility-based preferences for risk cannot explain conservative choice. Specifically, we show that adopting a model of expected utility (EU) under any plausible degree of risk aversion, r, modeled as any CARA utility in the interval, $r \in [0.0003, 0.005]$, does not increase the explanatory power of the benchmark. The upper bound of this interval indicates a degree of risk aversion so severe as to imply the rejection of a 50/50 gamble in which one either loses $175 (the 25th percentile of GQ rewards) or gains an infinite sum. While assuming (extreme) risk aversion moderately reduces the share of seemingly conservative choice (e.g., Goal 2 rather than Goal 3) relative to a risk-neutral benchmark, it increases the share of seemingly aggressive choice by roughly the same magnitude. Even a hyper-flexible benchmark of heterogeneous risk preferences, that classifies choice as explained if it can be rationalized by *any* risk preference within the interval, fails to explain 45 percent of choice.

Our third finding is that departures from the EU framework—biased beliefs, non-linear decision weights, and gain-loss utility—routinely contemplated as alternative explanations for risk aversion cannot explain conservative choice. For example, while a systematic bias in beliefs privileging lower goals (e.g.,

relative overconfidence about attaining lower goals) could, in theory, produce conservative goal choice among otherwise utility-maximizing employees, we document substantial employee overconfidence in both relative and absolute beliefs of attaining higher goals. Accordingly, a benchmark model of subjective expected utility (SEU), with plausible risk aversion, explains only one-half of employee choices, a rate only modestly improved after allowing for error in reported beliefs. Similarly, while conservative choice could result from the assumption of non-linear decision weights, adopting the widely-cited weighting function of Prelec (1998) did not improve the benchmark's explanatory power. Finally, given the precedent in the literature for explaining conservative choice through an aversion to (prospective) losses, we tested whether incorporating gain-loss utility into the benchmark would improve descriptive accuracy. We accomplished this by constructing a portfolio of benchmarks reflecting an exhaustive combination of loss aversion parameters, functional specifications, and candidate reference points informed by theory and the physical configuration of the menu. The 70 benchmarks we tested did not systematically explain a greater share of goal choice—the most successful among them accurately predicted 59 percent of choice.

To generate additional evidence on mechanisms and to rule out potential confounds, we designed and administered an experimental incentive-compatible rewards program, resembling GQ, in the context of an online effort task. The experimental paradigm permitted us to observe goal choice in a setting where we could confirm understanding of program rules, explicitly denominate rewards in dollars, rule out motives pertaining to reputation, signaling and costs of effort, and directly elicit/assess person-specific decision-making parameters. Because we elicited six goal choices per subject across strategically varying menus, the paradigm also afforded higher-powered tests of mechanisms. The exercise yielded a similar distribution of goal choice, overconfidence, and conservative choice (relative to the SEU benchmark) as in the field. And in evaluating choice relative to prior benchmarks, we found that none could explain all six decisions for more than one-quarter of participants. We additionally tested—and found no evidence to support—alternative explanations that we had been unable to test in the field such as heuristic strategies in which people sort into goals in an ordered menu based on relative self-assessments of ability or taste for competition (e.g., Kamenica 2009; Niederle and Vesterlund 2007). The evidence from the lab corroborates the serious limits to the descriptive accuracy of the EU framework indicated in the field.

We conclude by proposing a novel heuristic to help explain the observed pattern of risky choice in the present setting and perhaps more generally. Informed by open-ended descriptions of decision processes from pilot studies and our reading of the broader literature on inference and decision-making, our proposed "pairwise heuristic" stipulates that employees select a goal through a succession of approximate comparisons between pairs of proximal goals. Critically, the contingent inference prompted by the pairwise comparisons is subject to systematic bias, leading employees to underestimate the relative likelihood of attaining the riskier goal and increasing the likelihood of conservative goal choice. As a

concrete example, the heuristic implies an employee deciding between Goals 2 and 3 (having ruled out the Goal 1) would roughly assess whether the expected potential gain from selecting the high, relative to the low, goal (i.e., the difference in rewards weighted by the perceived conditional likelihood of high-goal attainment, $\Delta r_{3,2} * \hat{s}_{3|2}$), offsets the potential loss (i.e., $\hat{s}_{\neg 3|2} * r_2$). The employee, however, underestimates the conditional likelihood of goal attainment due to insufficient adjustment for the contingent nature of the comparison ($\hat{s}_{h|l} = ks_{h|l}$, $k \in [s_l, 1)$). Applied to GQ, the heuristic predicts greater variation, and conservatism, in goal choice than the prior benchmarks. While the proposed heuristic has not been previously discussed in the literature, it draws on well-established conjectures from the literature such as the propensity for relative evaluation (e.g., Bushong, Rabin, and Schwartzstein 2021), bias in contingent inference (see Benjamin 2019) and probability neglect (e.g., Rottenstreich and Kivetz 2006; Sunstein 2002), and noisy decision rules (e.g., Camerer 1989; Kahneman et al. 2021).

We sought evidence for the proposed heuristic from a new experiment. The experiment—which asked several hundred employees to make a hypothetical goal choice from a representative goal menu and then queried decision-making and beliefs—yielded several insights as to the plausibility of the heuristic. First, beyond corroborating the pattern of diverse and conservative (relative to the risk-neutral SEU benchmark) choice from the field, the experiment affirmed participant use of proximal pairwise comparisons and their systematic, and often substantial, underestimation of contingent likelihoods of goal attainment (participants exhibited a similar degree of bias in an unrelated domain). Second, conditioned on other decision-relevant beliefs, the magnitude of the within-subject bias strongly predicted optimal choice. Last, when randomized to select a goal from a "de-biased" menu—i.e., a menu displaying empirically-informed contingent likelihoods of goal attainment—participants were 48 percent more likely to select an optimal goal than a menu displaying the equivalent non-contingent likelihoods. Moreover, participant response to a menu displaying no probabilistic information about attainment (instead, we provided participants with a simulated performance history) was indistinguishable from response to a menu that displayed contingent likelihoods adjusted for the hypothesized bias ($k = s_l$).

As the final test of the heuristic, we compared its descriptive accuracy to prior benchmarks. With a conservative noise allowance, the pairwise heuristic explained 54 to 60 percent more choice than the SEV benchmark in the lab. In the field, despite an inability to observe the employee-specific bias in beliefs, a parametric representation of the heuristic explained 26 to 46 percent more choice than the same benchmark. In both the lab and field, the heuristic outperformed the other tested benchmarks. The evidence suggests, at a minimum, that a non-trivial share of employees relied on a decision-strategy resembling the proposed heuristic for goal choice with the likely possibility of far more widespread use.

Taken together, we see the present research as contributing to a growing literature seeking to understand financial risk-taking in the field in a setting of high transparency, generalizability, and

variability in stakes (see Barseghyan et al. 2018). While we find evidence for substantial risk aversion, we conclude that roughly one-half of employee decisions are inconsistent with expected-utility based motives. The evidence leads us to speculate that choice may arise from a novel choice heuristic in which the propensity to make relative comparisons across pairs of options induces biases, such as contingency neglect, often associated with relative inference. Beyond choice from goal-reward menus, we speculate that the proposed pairwise heuristic could help to explain risky choice in other contexts involving ordered-menus, such as portfolio allocation or insurance. To explore this possibility, we administered a final online experiment examining whether a heuristic involving contingency neglect could help explain excessive risk aversion and choice heterogeneity in the demand for home insurance. Using a stylized menu adapted from real-life data (Sydnor 2010), the experiment showed that participants making hypothetical insurance decisions from a menu encouraging non-contingent evaluation were substantially more prone to select the cheaper, and almost-certainly more valuable, plan than from either a baseline menu or a menu encouraging contingent evaluation. Collectively, the evidence emphasizes the importance of understanding the potentially non-standard decision-processes for the optimal design and unbiased welfare analysis of programs and policies involving financial risk.

2 BACKGROUND

2.1 Institutional Background

GoalQuest® (GQ) is an incentive program conceived and administered by BI WORLDWIDE (BIW), a private global consulting firm. The firm 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, channel partner, and consumer engagement. Founded in Minneapolis, Minnesota in 1950, the firm, as of 2021, had more than 25 sales offices across 9 countries and had self-reportedly engaged 6 million individuals across 144 countries through its various products. According to third-party estimates, as of 2021, the firm had approximately 1,500 employees and annual revenues between $500 million to $1 billion.

Described by BIW as the "world's only patented sales incentive design," GQ was designed to motivate employee productivity through self-selected performance goals tied to all-or-nothing non-monetary rewards.[3] 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

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

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 an impressive diversity of sectors (e.g., communications, health care, manufacturing, financials, consumer discretionary, consumer staples). Typically, a GQ program engages hundreds to a few thousand employees with the largest programs involving tens of thousands of employees. Employee participants range from newly hired to highly tenured and from staff workers to managers and supervisors.

2.2 Reward Program Overview

Across the wide-range of client firms, GQ boasts a uniform program structure, particularly since 2012, the year of our earliest data. From an employee's perspective, program participation entails three phases: enrollment/goal selection (and program marketing), a performance period typically lasting between 30 and 90 days, and, for those achieving their selected goal, reward redemption. During the initial phase, employees are asked to enroll in the program, and select their goal, by visiting an online portal and proceeding through a simple webflow.[4] The webflow itself consists of four phases: a program overview, an enumeration of program rules, goal selection, and goal confirmation (see Appendix for select screenshots). To select a goal, in most programs, employees are asked to select from a menu of three personalized goals (Goal 1, Goal 2, Goal 3) each associated with an all-or-nothing reward.[5] Rewards are denominated in points that employees can later redeem for prizes in a marketplace. BIW promotes the program as having a 98 or 99 percent participation rate among eligible employees. 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 the often-valuable rewards associated with goal attainment.

In 2014, we asked BIW to implement an enhanced enrollment process to elicit additional data from employees including their beliefs about goal attainment. Under enhanced enrollment, respondents were prompted to complete a brief survey immediately after selecting their goal. The survey asked employees to estimate their perceived likelihood of attaining each of the goals: "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?" (scale indexed in increments of 10 percentage points). Employees were additionally asked about their gender, age, and tenure with the firm. While the survey was optional, due to its integration within the enrollment process, survey participation across our sample was a robust 60 percent.

Following goal selection, employees transitioned to the several-week performance period during which they attempted to achieve their selected goal. In most programs, participants were able to log onto

[4] While the design of the enrollment portal was standardized across programs at any point in time, its design evolved over time.
[5] In some programs, the goal selection period may have briefly overlapped the performance window.

the website to check their progress or to remind themselves of their selected goal.[6] At the close of the performance period, employees who attained their goal could exchange their reward points for an actual reward in the GQ marketplace. The non-monetary rewards included major electronics (e.g., a 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; for many programs, employees were familiar with the marketplace through other BIW programs using the same point currency.

2.3 Goal and Reward Structure

Several elements of the GQ goal and reward structure were designed to increase employee productivity, premised on research in behavioral science.[7] First, based on the motivational power of self-selection and personalization (X), GQ asked employees to self-select a goal from a personalized menu. Specifically, excepting employees without any experience, a personalized goal menu was generated by applying a uniform rule to an employee's productivity during some baseline period prior to program administration (e.g., productivity during the prior quarter).[8] Almost all 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 $0.10x_b$). To further increase personalization, employees within a program were typically segregated into a small number of distinct groups based on factors such as their baseline performance, experience, or job type. While goal menus within each group were personalized using a single rule, rules could vary across groups. For example, this segregation strategy permitted GQ to assign new employees to a menu not informed by baseline data or to use different rules for employees who differed widely in their baseline performance. To ensure goals were genuinely self-selected, employees were not nudged towards a particular goal during the goal-selection process via recommendations, defaults, or persuasion.

Second, based on the presumed motivational power of goals, GQ encouraged higher goal choice by designing them to be more financially attractive, in expectation, for most employees. Specifically, in contrast to the typical menu with linearly increasing goals, rewards typically increased in non-linear increments. For example, many reward menus followed the k, $3k$, $6k$ structure, where k was set to be approximately 1 percent of an average employee's salary over the course of the program. Moreover, goals

[6] 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.
[7] Our insights into the origins of GQ draws from promotional materials, BIW white papers, and conversations with BIW leadership (e.g., see public-facing GoalQuest website, accessed December 2021).
[8] 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.

were all-or-nothing such that an employee selecting Goal 3 and achieving the Goal 2, but not Goal 3, threshold would earn no reward while an employee selecting Goal 1 and achieving the Goal 3 threshold would only earn the Goal 1 reward. As a result of reward non-linearity, and the all-or-nothing design, we estimate that, under rational expectations, Goal 3 maximized expected value for 84 percent of employees (Goal 2 maximized expected value for 11 percent of employees). Finally, the rewards associated with each goal were non-monetary, due to a belief that non-monetary rewards would be more motivating than equal-sized monetary rewards.

3 THEORETICAL FRAMEWORK OF GOAL CHOICE

We now introduce a theoretical framework to organize our analysis of conservative goal choice. We represent the GoalQuest program as a choice between two simple lotteries and assume employees select the goal that maximizes their expected utility given their beliefs of goal attainment. We then amend the model to consider systematic departures from the standard framework such as the potential for biased beliefs, non-standard decision-weights, and reference-dependent utility. Finally, we consider the possibility that conservative choice can be attributed to heterogeneity in decision-making frameworks.

3.1 Generalized Expected Utility Framework

We begin by outlining a generalized framework to describe how a utility-maximizing employee selects a productivity goal associated with an all-or-nothing reward from a menu of choices. To simplify matters, we model the menu as having just two options, a high goal (higher difficulty, higher reward) and a low goal (lower difficulty, lower reward). Employees understand that, in the period following goal selection they will earn the pre-specified reward if their level of productivity meets or exceeds their selected goal. We assume that the level of realized output is attributable in part to employee-specific ability, non-employee-specific productivity shocks, but does not depend on goal selection.

Formally, we represent goal choice as participation in one of two available lotteries, $G_n \in [G_h, G_l]$. Each lottery yields a reward x_n with some probability s_n and 0 with some probability $(1 - s_n)$. In our context, the high goal has a strictly higher reward and lower likelihood of attainment than the low goal, $x_h > x_l$ and $s_h < s_l$. Subjective probabilities, $\hat{s}_n$, capture employee beliefs about actual likelihoods of attainment. An employee must select exactly one of the two lotteries. We can describe an employee's valuation of a goal-lottery as follows: $\mathrm{V}(G_n) = \pi_n(\hat{s}_n)\, u(x_n, \theta)$. Here, π_n denotes the decision weight an employee assigns to the subjective probability of goal attainment, and $u(.)$ is an always increasing function that captures an employee's preference rewards and potentially depends on a reference point, θ.

An employee will choose the low goal if $\pi_l v(x_l, .) \geq \pi_h v(x_h, .)$. Our generalized framework identifies several potential reasons for why an employee might choose a conservative goal including a

substantial difference in expected likelihood of attainment, the curvature of their utility function, biased beliefs about goal attainment ($\hat{s}_n \neq s_n$), and non-linear decision weights ($\pi(\widehat{s_n}) \neq \widehat{s_n}$).

3.2 Conservatism with EUT and Rational Expectations [$\pi(s) = s$, $v(x_n, .) = u(x_n)$]

As a baseline informed by expected utility theory, a well-informed employee selects the goal that maximizes expected utility using linear decision weights, a utility function dependent only on final wealth states, and rational expectations of attainment. For tractability, we assume a parametric utility function from the constant absolute risk aversion (CARA) family, so that the parameter, *r*, captures an employee's attitude towards risk, *r* > 0 implies risk aversion, and *r* = 0 denotes risk neutrality (we ignore the possibility that *r* < 0):

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

While our choice of a CARA function permits us to represent risk attitudes with a single parameter, it implies the irrelevance of an employee's prior wealth for risk preferences. We speculate that abstracting away from initial wealth is reasonable given that an employee must evaluate two lotteries relative to a single level of initial wealth. Nevertheless, we contemplate utility functions featuring constant relative risk aversion for varying levels of initial wealth in the Appendix. Our assumption of rational expectations implies that employees have unbiased and well-informed beliefs regarding the likelihood of goal attainment, $\hat{s}_n^r$, such that $\hat{s}_n^r = E(s_n \mid \Phi) = s_n + \varepsilon$. Here, Φ is the information set available to an employee at the time of goal choice and ε is a normally distributed, mean-zero, error term with constant variance.

<u>Risk Neutrality (*r* = 0)</u>. For completeness, we first consider the case of risk neutrality. An employee who is indifferent to financial risk will choose the low goal if: $\hat{s}_l^r / \hat{s}_h^r > x_h / x_l$. Given these preferences and beliefs, we should expect to observe conservative goal choice only if the relative expected likelihood of achieving the low versus high goal exceeds the ratio of the high versus low goal.

<u>Risk Aversion (*r* > 0)</u>. Next we consider the more plausible scenario in which an employee is averse to financial risks. Such an employee will choose the 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 that conservative goal choice is positively increasing in the degree of risk aversion, as well as expectations of relative goal attainment, and the gap between high and low goal rewards. We consider risk aversion parameters within some range of plausibility $r < r'$. Practically, we

establish an upper bound of plausibility by examining the behavior implied by such risk preferences in simple lotteries involving financial stakes comparable to those engaged in the GQ program.

Heterogeneous Risk ($r_i \in [0, r']$). Finally, we consider the possibility that employees exhibit heterogeneity across their risk preferences. The possibility of non-uniform attitudes towards risk has been explored in prior research (X). We specifically consider whether the goal choices of employees can be rationalized by any degree of risk aversion within an interval of plausibility, $r_i \in [0, r']$.

3.3 Conservatism Due to Non-Standard Beliefs [$\hat{s}_n \neq E(s_n \mid \Phi)$]

We next consider the possibility that conservative goal choice emerges from the non-standard beliefs of a risk averse employee who maximizes an expected utility function with linear decision weights. We can depict non-standard beliefs with a multiplicative constant, $\hat{s}_n = \gamma_n s_n + \varepsilon$, such that γ_n represents the degree of goal-specific distortion to beliefs. As a result, $\gamma_n > 1$ implies overconfidence while $\gamma_n < 1$ implies underconfidence.

A risk averse employee with distorted choices of this nature will choose 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 that conservative goal choice increases in γ_l/γ_h. If $\gamma_l/\gamma_h > 1$, the share of conservative choice should be higher than implied by the standard benchmark. If $\gamma_l/\gamma_h < 1$, the share of conservative choice should be lower than implied by the benchmark. If distortion of beliefs is symmetric across the low and high goals, the share of conservative choice should not differ from the benchmark.

3.4 Conservatism Due to Non-Standard Decision weights [$\pi(s) \neq s$]

We now assess whether the adoption of non-linear decision weights helps to explain employee behavior. In particular, researchers have advanced several probability weighting functions to address violations of expected utility in which people appear to overweight highly improbable outcomes and underweight highly probably outcomes. Given the literature's emphasis on an inverse-S shaped weighting functions, we adopt arguably the most popular of these functions, the function proposed by Prelec (1998):

$$\pi_n = \exp(-(-\ln s_n)^{\alpha})$$

The non-linear weighting function could result in conservative goal choice if an employee were to underweight the likelihood of attaining a high goal relative to attaining a low goal. The decision rule for an employee governed by non-linear decision weights is given by:

$$r > \frac{\ln\left(\frac{\pi_l(s_l)}{\pi_h(s_h)}\right)}{x_l - x_h}$$

3.5 Conservatism Due to Loss Aversion $[u(.) = u(x_n, \theta)]$

Finally, we consider the possibility that conservative goal choices may arise as the result of employees exhibiting loss aversion in the context of gain-loss preferences (Kahneman and Tversky 1979; 1992). Loss aversion has been advanced as a possible theoretical explanation for small- to moderate- scale risk aversion by Rabin (2000) and Rabin and Thaler (2001) and has practically been suggested as an explanation for field evidence in contexts from insurance (Sydnor 2010), investments (X), and betting (X). Given that the structure of GQ stipulates that every employee receives either nothing a positive reward, employees do not engage explicit losses in the program context. However, following the expectation-based approach of Koszegi and Rabin (2006), it is reasonable to interpret goals, especially those associated with substantial rewards, as potential reference points (Heath, Larrick, and Wu 1999).

Specifically, given some reference point, θ, we can describe utility as follows:

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

As is standard practice, we specify that utility over gains, u^+, is concave, while utility over losses, u^- is convex. While the literature does not provide clarity as to how practically specify a reference point in the context of the goal menu, in the empirical analysis we consider a range of plausible reference points informed by both theoretical and exploratory work.

4 DATA AND SAMPLE CONSTRUCTION

Our analysis of financial decisions under risk leverages program- and employee-level administrative data from BIW. The employee- level data included demographic detail, goal choice, employee productivity over the duration of the program, and employee beliefs elicited from enhanced enrollment. The program-level data identified each firm (and department), the date of program administration, rules used to segregate employees into groups, and details of the goal/reward menus faced by each participant. In this section, we describe the construction of our analytic sample, summarize its key features, and define the variables central to the subsequent analysis.

4.1 Primary Sample

Our analysis of employee behavior primarily engaged a *primary sample* that we constructed by applying screening restrictions to an original dataset from BIW. This original data, which spanned 38,661

employees across 34 programs and 18 firms, reflected the universe of data from GQ programs administered between 2014 to 2018 in the US or Canada that had adopted enhanced enrollment, had at least 100 participants who completed the program, and whose data had been electronically archived by BIW.[9] As an intermediate step towards producing the primary sample, we generated an *expansive sample* by excluding about 8 percent of records from the original data for which a key data field was missing (excluding employee salary for which we only have partial coverage), the data was inconsistent (e.g., data on performance didn't match goal choice and received reward), or we inferred the employee was not likely to have completed the program (n = 35,478).[10] We then restricted the expansive sample to those who completed enhanced enrollment and provided internally consistent beliefs, resulting in a primary sample of 20,133 decisions and corresponding beliefs.[11] In comparing the expansive and primary samples, we find that employees completing enhanced enrollment were moderately more likely to select aggressive goals and modestly more likely to attain them, implying that the conservatism and sub-optimal choice that we subsequently document in the main analysis may slightly underestimate the degree of conservatism and sub-optimal choice in the broader population of employees.[12] In light of potential selection into enhanced enrollment, we reproduce key analyses for the expansive sample in Section 5.

The primary sample comprises the decisions of 20,133 employees across 18 firms, 34 programs, and 232 distinct groups.[13] Table 1 summarizes overall sample statistics as well as group-level (duration, financial stakes) and employee-level (age, gender, tenure, inferred income) characteristics. On average, we observe data for 592 employee participants per program (IQR: 208 to 703) and 87 employees (IQR: 12 to 103) per group. Groups varied roughly uniformly across three categories of duration, 30, 60, and 90 days (there are two outlier programs that ran for 45 and 120 days). While the group-level average (median) potential reward was $607 ($375), the distribution of potential reward values was asymmetric, such that 10 percent of employees faced decisions involving rewards worth an average of $2,150. Overall, employees in the sample could have earned up to $9.4 million in possible rewards.

---

[9] Data for a small number of programs was not archived by BIW. The program size cutoff was practically necessitated by the administrative burdens of organizing and transferring program data (BIW) and the resources required by our team to process and manually audit data from each program.

[10] Approximately 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.

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

[12] 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).

[13] Some firms participated in multiple GoalQuest programs sequentially at the same location, so a small number of employees appear in the sample multiple times in different programs.

The table also indicates the diversity of the employee sample with respect to gender, age, and tenure. We speculate that the high share of employees, 73 percent, with tenure of 5 years or less likely reflects the fact that GQ programs are often administered in industry sectors with high turnover. While we only observe average program-level salaries for programs in 8 of the 18 firms, or 25 percent of the sample, the average of this program-level statistic is $70,400, well above the average U.S. individual income. However, because the availability of salary data is greater for programs with higher rewards, we suspect the average salary among employees in our sample is closer to the national average.

4.2 Goal Choice and Employee Productivity

We turn next to data on goal choice and employee productivity. We describe goal choice, *g*, both through indicators for goal choice as well as indicators denoting whether the goal reflects an optimal, aggressive, or conservative choice. The characterization of goal choice relies on a comparison of the expected utility associated with the selected, and non-selected, goals with respect to the benchmark models outlined in the theoretical framework. We describe employee productivity through two normalized measures: (i) productivity relative to the baseline threshold and (ii) productivity relative to the Goal 3 threshold.[14] Normalization facilitates comparisons across programs whose productivity outcomes may vary substantially in their scale (e.g., at a call center, productivity may be measured in hundreds of calls over the program period, while at a sales office, it may be measured in single-digit unit sales). Finally, we report indicators of baseline and goal attainment for each employee.

Table 2 summarizes the choice, productivity, and attainment measures. The table indicates that 44 percent of employees selected the highest goal while remaining employees roughly split across a choice of the low and medium goals. The performance of the median employee nominally exceeded the baseline threshold (1.01 ratio) and moderately fell short of the Goal 3 threshold (0.89 ratio). Said differently, while just over one half of employees reached their baseline threshold, only 29 percent reached the Goal 3 threshold. The table also indicates a correlation between goal choice and productivity—for example, 40 percent of those selecting Goal 3 attained the goal—suggesting that more productive employees select into higher goals (or possibly that higher goal choice led to heightened performance).

4.3 Employee Beliefs

Our analysis draws on two measures of employee beliefs—raw subjective beliefs elicited through enhanced enrollment and estimates of ex ante rational expectations. To construct a measure of subjective

[14] We do not have employee baseline data for 16% of the sample. In some cases, this reflects the lack of past performance data for new employees. In other cases, performance goals were defined without reference to an employee's past performance and no baseline data was shared with BIW. The average performance relative to baseline is calculated on the remaining fraction of the sample for which baseline data is available. See Background section for more information on the origin of the baseline data.

expectations describing employee i's probabilistic belief, $\hat{s}_{k,i}$, of attaining goal threshold k, we simply record the employee's response in enhanced enrollment, in 10-unit increments on a scale from 0 to 100 percent. Our estimation of an employee's ex ante rational expectation of goal attainment, $\hat{s}^{r}_{k,i}$, consists of two steps. First, we calculate the average ex post goal attainment for each program group and each goal. Next we predict the ex ante likelihood of goal attainment for each goal and each employee by adjusting the group-average by observable covariates. The exercise effectively assumes that one can proxy for rational expectations by appealing to average attainment for similar others. The estimation strategy is consistent with the intent of program administrators to sort employees with similar performance expectations (or, in the case of new employees, a similar lack of baseline data) in the same group.

To implement the strategy, we first 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 $\mathbf{Z}$ (age, tenure, gender) and group fixed effects, π_l. We then calculate an employee's rational expectation of obtaining goal k, as $\hat{s}^{r}_{k,i} = \hat{\alpha} + \boldsymbol{Z}\hat{\gamma} + \hat{\pi}$. We estimate regressions at the program-level so that we have a larger sample from which to estimate employee covariates. This technique of inferring rational expectations by looking at the realized outcomes of similar others is common in economic analyses of insurance choice (e.g., Bhargava et al. 2017).

Table 3, which summarizes belief data, indicates two patterns of interest. First, the table suggests that increasing expectations of attainment predicted goal choice. Employees selecting more aggressive goals had, on average, higher (rational and subjective) expectations of attaining those goals. For example, 43 percent of those selecting Goal 1 subjectively expected to attain Goal 3, compared to 76 percent of those selecting Goal 3. Second, the comparison between subjective beliefs and our estimates of rational expectations suggests that employees were substantially overconfident, on average, about their future productivity. We investigate employee overconfidence in greater depth in the next section.

5 CHARACTERIZATION OF GOAL CHOICE BY BENCHMARK MODEL

We now turn to characterizing employee choice relative to predictions of the benchmark models outlined in the theoretical framework. For each benchmark, we report the share of conservative, optimal, and, for completeness, aggressive, choice. To better understand the moderating role of financial stakes and experience, we additionally report optimal choice shares across reward size and employee tenure.

5.1 Expected Utility with Risk Neutrality

We begin by assessing choice relative to a baseline benchmark which presumes that risk-neutral employees choose a goal that maximizes their expected utility given beliefs of goal attainment. We

initially assume perfect information before considering the case of rational expectations and subjective beliefs, directly elicited via enhanced enrollment. For each benchmark, we report the share of goal choices that match the prediction of the model (optimal choice), the share of goal choices that involve a goal lower than the model prediction (conservative choice), and the share of goal choices that involve a goal higher than that predicted by the model (aggressive choice).

Perfect Information Benchmark. We first consider the scenario in which employees have perfect information regarding their eventual productivity. The unrealistic, but instructive, exercise is equivalent to characterizing choice based on an employee's observed ex post productivity, assuming that productivity is unaffected by goal choice (an assumption we subsequently revisit). We restrict this analysis to the 44 percent of employees who attained Goal 1 since, for remaining employees, the ex post benchmark implies the irrelevancy of choice. The first panel of Table 4 characterizes goal choice for the baseline benchmark model across different assumptions regarding beliefs. The table indicates that, assuming perfect information, just over one-half of employees chose optimally, while 31 percent of employees chose conservatively. For employees who attained Goal 1, conservative choice resulted in an average unrealized reward of $272 (IQR of $63 to $400), where unrealized refers to the difference between the realized reward and the counterfactual reward one would have achieved under optimal choice. Unrealized rewards associated with conservative choice amounted to 58 percent of the optimal reward. Figure 1, which depicts the cumulative distribution of unrealized rewards by goal choice, for employees achieving at least Goal 1, indicates that most unrealized rewards are concentrated among employees who selected either Goal 1 or Goal 2. In this sense, the figure confirms the optimality, under perfect information, of Goal 3 for most employees. Finally, to better understand the role of economic moderators, the second panel of the table reports how the optimality of choice varies across the potential reward stakes (indexed by quartile) and employee tenure (indexed by response category). The table implies that the share of optimal choice did not substantially increase with higher potential rewards or increased employee experience.

Rational Expectations. We turn next to characterizing goal choice under the assumption that employees have rational expectations. For this, and subsequent, benchmarks, we abandon the earlier and now unnecessary sample restriction of Goal 1 attainment. The table indicates that, relative to perfect information, rational expectations shifts the characterization of choice towards increased conservatism and reduced optimality (0.45 optimal; 0.49 conservative; 0.06 aggressive). We note, however, that the comparison reflects the effects of both a change in the information assumption and compositional differences emerging from the relaxation of the sample restriction. For a clearer understanding of the former, we calculate that the transition from perfect information to rational expectations in the restricted sample would shift characterization from (0.51 optimal; 0.31 conservative; 0.18 aggressive) to (0.52 optimal; 0.43 conservative; 0.05 aggressive), suggesting that the adoption of rational expectations results

in an increase in conservative choice share. The intuition for the shift is that rational expectations would interpret the choices of many employees who selected Goal 2 as conservative, even if they did not achieve the goal. This is due to the overwhelming attractiveness of Goal 3 in expectation. The table also indicates that the adoption of rational expectations, relative to perfect information, results in a reduction in the magnitude of the average unrealized reward. Much of this reduction, however, is mechanical, since the rational expectations benchmark weighs rewards by the expected likelihood of attainment. As with perfect information, the share of optimal choice under rational expectations is not moderated by reward size or employee experience. Finally, we note that characterizing choice in the expansive sample (i.e., the sample inclusive of the primary sample and employees not completing enhanced enrollment) results in an approximately similar distribution of choice (0.43 optimal; 0.53 conservative; 0.04 aggressive).[15]

Subjective Beliefs. We proceed to consider the possibility that the high share of apparent conservatism implied by the rational expectation benchmark may reflect systematic bias in employee forecasts of productivity. Such bias could account for conservative choice if employees held inflated beliefs about their relative likelihood of attaining low versus high goals. For example, if employees were systematically under-confident about future productivity, and this led employees to inflate the likelihood of achieving lower, relative to higher goals, then one might expect utility-maximizing employees to select lower goals more frequently than predicted by the benchmark model. Alternatively, employee overconfidence could similarly explain the high share of ostensibly conservative choice if such overconfidence led to employees to systematically inflate the likelihood of achieving lower, relative to higher, goals. To assess whether the observed choice patterns reflect biased beliefs, we re-characterized goal choice after replacing rational expectations of goal attainment with subjective beliefs elicited from the enhanced enrollment process. The results, reported in Table 4, indicate that adopting subjective, rather than rational, expectations, did not reduce the share of conservative choice and led to only modest improvement to overall descriptive accuracy. As with the prior benchmarks, reward size and employee experience did not predict more efficient choice in the subjective utility benchmark.

Table 3 provides additional insight into how employee beliefs affect the characterization of choice. For each goal, the table reports the average rational and subjective beliefs of attainment and the average ratio of subjective and rational expectations for the entire sample and separately by goal choice. We highlight three patterns of note from the table. First, Panel A indicates that subjective beliefs of goal attainment strongly predict goal choice, suggesting that the elicitation produced credible results. Second,

[15] To characterize choice under rational expectations in the expansive sample, we follow the same strategy described above but for a simplified estimation of rational expectations. Specifically, to accommodate the absence of demographic data among those not completing enhanced enrollment, our regression estimates of rational expectations exclude covariates beyond the group-level indicator (effectively, such estimates are equivalent to the leave-out group-level mean).

Panel B indicates that subjective beliefs of attainment substantially exceed rational expectations for each of the three goals by a ratio ranging from 2.20 (Goal 1) to 3.46 (Goal 3). This suggests that employees exhibit substantial overconfidence across all goals. In visually comparing the distribution of rational and subjective beliefs for each goal, Figure 3 corroborates substantial overconfidence for each goal. Finally, the ratios of relative overconfidence reported in the second panel of the table suggest employees are, on average, more overconfident about attaining higher, relative to lower, goals. Ultimately, the table and figure emphasize that biased employee beliefs cannot explain conservative goal choice, as employees are both absolutely, and relatively, overconfident about increasingly higher goal attainment.

5.2 Expected Utility with Risk Aversion

We turn now to the possibility that conservative choice may reflect risk aversion attributable to the diminishing marginal utility of wealth. As described in the theoretical framework, we model risk aversion by assuming a utility function that exhibits constant absolute risk aversion governed by parameter r. We represent the plausible range of risk aversion with the interval, $r \in [0.0003, 0.005]$. To appreciate the breadth of risk attitudes captured by this interval, one can translate what such risk preferences imply for gambles involving potential losses of a size similar to the potential rewards engaged by employees in GQ. For example, consider a simple lottery involving a 50 percent chance of losing $175 (the 25th percentile GQ reward value) and a 50 percent chance of some unspecified gain. A risk aversion parameter of r = 0.0003 implies that an employee would accept any such gamble so long as the potential gain exceeds $184—a modest, but seemingly plausible, degree of risk aversion. The same employee should accept any 50/50 gamble involving a potential loss of $350 (the median GQ reward value) so long as the potential gain exceeds $391. At the other endpoint, r = 0.005 implies that an employee would reject *any* 50/50 gamble involving a potential loss of $175 (or $350), even if the potential gain was infinite. We therefore treat r = 0.005 as a highly conservative upper bound of plausible risk aversion for financial stakes in the range of interest.

Table 3 characterizes choice for risk-averse, expected utility maximizing, employees under either rational or subjective expectations. Perhaps unsurprisingly, the assumption of modest risk aversion (r = 0.0003) does little to shift the characterization of choice, across either information regime, relative to risk neutrality. However, across both information regimes, the assumption of severe risk aversion (r = 0.005) shifts the characterization by moderately reducing the apparent share of conservative choice, largely offset by an increase in the share of choice characterized as aggressive. Incorporating risk aversion into the benchmark model does not substantially shift descriptive accuracy (i.e., the share of optimal choice). Risk aversion also does not affect the absence of moderation by reward size or experience.

Figure 4 conveys the intuition for the shift in choice characterization under risk-averse preferences. The figure depicts the share of employees whose goal choice is deemed as optimal by the benchmark model overall, and by goal choice, assuming rational (Panel A) and subjective (Panel B) beliefs, for r ranging from $r = 0$ to $r = 0.10$. Across panels, the figure shows that as one increases the degree of assumed risk aversion within the plausible range (highlighted region), the share of optimal choice among employees choosing Goals 1 and 2 increases moderately, but this increase is offset by the reduced share of optimality among employees choosing Goal 3. Seen another way, greater risk aversion reduces the share of conservative choice among low-goal choosers and increases the share of aggressive choice among Goal 3 choosers. Ultimately, the exhibits suggest that the incorporation of plausible risk aversion, induced from the concavity of the utility function, does not improve the descriptive fit of the benchmark models and does not explain the conservatism of nearly 40 percent of employees.

Heterogeneous Risk Preferences. Conceivably, goal choice may reflect the utility-maximizing behavior of a population with heterogeneous risk preferences. To evaluate this possibility, we reassessed the optimality of choice after classifying any goal choice as optimal if it could be rationalized by *any* value of r within the interval $[0, 0.005]$. As reported in Table 4, shifting from a benchmark model assuming extreme, but uniform, risk aversion, to one assuming heterogeneous risk preferences increases the share of optimal goal choice from 0.44 to 0.55 percent (rational expectations) and from 0.53 to 0.59 percent (subjective beliefs). Allowing for highly flexible risk preferences also serves to increase the differential share of optimal choice across high and low reward size but not high and low employee experience. We note that the differential rate of optimal choice across reward size may be mechanical, since for menus with higher rewards, the likelihood that flexible risk preferences rationalizes the choice of two of the three goals on the menu is higher than it is for menus with smaller rewards. We revisit the possibility that decisions may reflect diversity in risk preferences in subsequent experimental analyses.

5.3 Non-Linear Decision Weights

We proceed to consider whether two commonly invoked behavioral departures from the standard expected utility framework can help to explain employee choice. The first departure we consider is the assumption of non-linear decision weights (i.e., a model of rank-dependent utility). Specifically, we consider a weighting function suggested by Prelec (1998; $\alpha = \beta = 0.65$) with an inverse s-shaped form commonly asserted in the literature. In theory, if employees systematically underweight moderate-probability outcomes (e.g., Goal 3) relative to higher-probability outcomes (e.g., Goals 1 and 2), the assumption of a non-linear weighting function might help explain goal choice.

Table 5 contrasts the characterization of choice assuming non-linear decision weights relative to a baseline model of subjective expected utility with modest CARA risk aversion, $r = 0.0003$ (reported in

the first column of Table 5 and reproduced from Table 3). The exercise indicates that non-linear decision weights do not meaningfully shift the characterization of choice relative to the subjective utility baseline. Given the modest transformation of beliefs entailed by typical non-linear weighting functions, the absence of a strong shift in characterization is perhaps unsurprising. Non-linear decision weights also do not change the previously documented absence of moderation across reward size or employee experience.

5.4 Gain-Loss Utility

Finally, we consider the possibility that conservative employee choice may reflect loss aversion in the context of gain-loss utility. While our context involves no explicit loss, research has suggested that goals may serve as reference points that invoke behavioral response consistent with those described in gain-loss paradigms (Heath, Larrick, and Wu 1999). One practical challenge for assessing models of gain-loss utility, however, is the absence of clear theoretical guidance as to functional form, magnitude of loss aversion, and the specification of the reference point. Regarding the latter, while Kahneman and Tversky (1979) originally adopted the standard quo as a reference point, they contemplated the potential for other reference points. Subsequent work has suggested a range of candidate reference points including those that are prospect-specific, expectation-based (Koszegi and Rabin 2006; Loomes and Sugden 1986), and/or informed by salient considerations such as the certainty equivalence of a gamble (Gul 1991) or features of the choice menu. Perhaps a more practical resource for identifying reference points is provided by Baillon, Bleichrodt, Spinu (2020) who evaluate the success of gain-loss utility models across potential prospect-independent (e.g., status quo, the high outcome, the highest probability option, the highest option a person is certain to achieve) and prospect-dependent (e.g., the selected option, the expected value of the selected option) reference points in explaining risky choice from choice-menus in the lab.

In light of our broad interest in all credible formulations of gain-loss utility, we aspired to characterize goal choice for a wide-range of plausible reference points informed by the theoretical and empirical literature as well as the configuration of the GQ menu. We consequently considered five prospect-independent reference points: status quo (i.e., $0), the high probability goal (Goal 1), the high reward goal (Goal 3), the highest goal an employee felt certain to achieve (otherwise $0), and, for completeness, Goal 2. We also considered prospect-dependent reference points including the chosen goal, the expected value of the chosen goal, and in light of the potential role of counterfactual regret, the proximal goal either below or above the chosen goal.

We assess these reference points in the context of two prominent functional representations of gain-loss utility—gain-loss utility in isolation, using the KT (1979) power function (initially setting α = 0.88) and a composite framework in which utility is comprised of both consumption utility and a gain-loss component (Sugden 2003; Kobberling and Wakker 2005; Koszegi and Rabin 2006, 2007). For the

composite functions, we assume that consumption utility and gain-loss utility are additively separable, adopt the same KT power function for both utility components, and denote η as a scaling factor applied to consumption utility such that setting $\eta = 0$ reduces to a model with gain-loss utility only. In deference to the range of loss aversion parameters contemplated by the literature, and the lack of clarity as to the appropriate weighting across the two components of utility, we each model for values of λ from 1.5 to 3.0 and η from 1 to 5. Finally, to streamline the analysis, we ignore probability-weights across all gain-loss models and evaluate each model assuming subjective expectations.

Table A1 of the appendix reports the descriptive accuracy of the gain-loss utility candidates. Among the prospect-independent reference points, nearly all of the reference points explain approximately one-half of all goal choices. To understand the intuition for this consistency, consider that for many employees, adopting a reference point shifts the utility-maximizing goal from Goal 3 to either Goal 1 or Goal 2. The net result is to increase optimality among low-goal choosers by roughly the same magnitude as the decrease in optimality among employees selecting Goal 3. Among prospect-dependent reference points, the chosen goal provides the most descriptively successful reference point while across the composite utility formulations, the reference point explains between 54 and 59 percent of choices.

Table 5 characterizes choice for the most promising of the gain-loss formulations. In deference to the prominence of models of regret in the literature, the table also reports the characterization associated with the counterfactual regret model. The table indicates that the incorporation of loss aversion via gain-loss utility delivers, at most, a modest increase in explanatory power relative to baseline. Once again, the table indicates that neither reward size nor employee experience moderate the descriptive accuracy of the models. Ultimately, our attempts to characterize choice with a generalized expected utility framework did not yield a model with rates of descriptive accuracy substantially exceeding 50 percent. Moreover, most models imply a high degree of conservative choice. We explore additional explanations, and attempt to rule out potential confounds, for the observed conservatism through an online goal choice paradigm.

6 EXPLORING MECHANISMS VIA ONLINE EXPERIMENTS

We further investigate the motives for conservative goal choice through two online experiments. The first study was intended to corroborate findings from the field with increased statistical power, assess alternative explanations from the literature, and rule out potential confounds. The second study was designed to test a novel explanation of GQ goal choice informed by exploratory pilot studies and the broader literature on decision-making and inference.

6.1 Online Goal-Reward Paradigm (Experiment A)

Overview. We administered the first experiment (Experiment A) in May 2019 on the Qualtrics platform to 407 employed US adults recruited from Amazon Mechanical Turk. The online instrument

asked participants to complete an effort task in the context of an incentive-compatible goal-reward paradigm. The paradigm resembled GQ but with lower stakes, a shorter evaluation period, comprehension checks, and multiple decisions per subject. We supplemented the goal-reward paradigm with several decision-relevant questions including an elicitation of beliefs, an assessment of risk and loss aversion through hypothetical gambles, and self-assessments of relative ability and taste for competition.[16]

Implementation of Goal-Reward Paradigm. We implemented the goal-reward paradigm by first explaining to participants that they would be partaking in a timed effort task where they could earn financial rewards for solving a series of grids. The effort task, which resembled those used in the literature, required participants to "solve" a 3 x 3 grid of single-digit numbers by finding the unique pair of numbers whose sum equaled 10. After an opportunity to practice, we formally introduced participants to the goal-reward paradigm, which we named GoalQuest, via a webflow resembling that used in field. The webflow explained participants would have four minutes to solve as many grids as they were able and that they could earn rewards by attaining self-selected performance goals from an all-or-nothing menu. After questions to ensure comprehension of the paradigm, participants proceeded to goal selection.

To increase statistical power, we asked participants to select a goal from each of six distinct menus, explaining one menu would be randomly selected to determine the participant's actual reward. The menus strategically varied the spacing of the goals and rewards as well as the number of options in order to facilitate tests of mechanisms (as informed by pilot tests of the paradigm). Specifically, we designated a baseline menu that resembled the field in additively linear goals (6, 8, 10) and non-linearly increasing rewards ($0.10, $0.20, $0.35). Four menus varied either overall difficulty or the relative attractiveness of Goal 3 and two menus expanded the choice-set by adding a relatively unattractive high- or low-goal option to the baseline offering. After goal selection, we elicited performance expectations by asking participants to estimate their likelihood of achieving various grid-thresholds. We used these expectations to impute beliefs for every goal across the six menus (a departure from the field where we could directly elicit beliefs of attaining each goal on the menu).[17] We additionally asked participants to forecast how many grids they would complete. Finally, participants completed the four-minute effort task.

---

[16] To assess relative grid-solving ability, we asked participants to evaluate their ability to solve grids relative to others the study on a five-point scale ranging from well-below average to well-above average. We elicited participants' self-perceived relative taste for competition on a five-point scale from "much less" to "much more" competitive than others.

[17] We impute expected performance as follows. First, we calculate the difference in subjective likelihood of completing n and $n+2$ grids, and assume the participant will complete exactly $n+1$ grids with this likelihood. For example, if a participant reports a 60 percent likelihood of completing 8 grids and 90 percent likelihood of completing 6 grids, we impute an expectation they will complete exactly 7 grids with 20% likelihood. To address expectations about performance below 4 grids and above 18 grids, we take the observed average performance among participants who complete less than 4 grids and more than 18 grids. On average, these participants complete 1.09 and 19.89 grids respectively. If participants assign any likelihood to completing less than 4 grids or more than 18 (that is, if their subjective likelihood of completing 4 grids is less than 100%, or their subjective likelihood of completing 18 grids is greater than 0%), we impute these values as the expected conditional performance. Finally, we sum these expectations across the entire distribution for each participant to arrive at a total expected performance.

Results – Comparison of Lab and Field. Discarding data from participants with incomplete or internally inconsistent beliefs resulted in a final sample of 277 participants who made 1,662 goal choices. Participants engaged the baseline menu in a manner similar to typical employees in the field. Average choice across the three goals (Goal 1, 2, 3) in the baseline men from the lab (0.34, 0.28, 0.38) roughly resembled choice in the field (0.29, 0.27, 0.44) as did beliefs of goal attainment in the lab (0.80, 0.66, 0.51) and field (0.78, 0.69, 0.63). Participants also exhibited overconfidence on average for each goal, but not as severely the field due to significantly higher goal attainment (overconfidence in the lab was higher for more challenging menus). Under the risk-neutral SEU benchmark, choice characterization (optimal, aggressive) was similar across the baseline condition in the lab (0.50, 0.45) and the field (0.50, 0.48).

The correspondence in choice, beliefs, and choice efficiency across the lab and field also discounts the possibility that patterns from the field were due to confounds involving program confusion, managerial signaling, or reputational concerns (while the latter two motives would presumably nudge employees towards more aggressive goals, in theory they could encourage more attainable goal choice). Notably, this correspondence was achieved in a setting with dollar-denominated rewards, verified comprehension of the paradigm, and no scope for signaling or reputational concerns.[18] We summarize choice, beliefs, and attainment for 3-option menus in the Appendix.

Results - Characterization of Goal Choice. Table 6 characterizes the optimality of choice relative to a range of benchmark models.[19] The characterization reported in the table reflects greater statistical clarity than the field because we observe multiple goal choices for each participant. The table also characterizes choice after allowing for some computational error in the form of a participants whose choices mostly, if not entirely, adhered to the benchmark. The first row of the table indicates that prior benchmarks can fully explain at most 24 percent of choices. Allowing for error, in the form of at least 5 of 6 optimal choices increases the explanatory upper bound to 42 percent of decisions. Notably, the rational expectation and subjective EU benchmarks fully explained only a modest share of participant decisions.

The table also reports tests of two additional heuristic choice-strategies, informed by the literature, that we were not able to test in the field. The first, contextual sorting, presumes that employees heuristically selected the goal whose relative position in the ordered-menu corresponds to their perceived standing in some choice-relevant distribution such as ability or productivity. This heuristic would be a sensible strategy for someone who was unsure of what goal to select but believed that the menu was

[18] An intriguing possibility is that conservative goal choice in the field largely reflects employee intent to pre-emptively commit themselves to easier goals so as to avoid the perceived effort costs associated with ambitious goals. While possible, we see this explanation as unlikely given (a) many conservative choices were made by employees who perceived high-goal attainment as very likely and (b) we find a similar pattern of conservative choice in the lab where effort-motives should be minimized.

[19] We assign rational expectations for each participant as the predicted likelihood of goal attainment estimated from a regression of goal attainment on observable characteristics and practice round performance across the sample.

designed so that each goal was optimal for a roughly equal share of participants. Contextual sorting of this sort was suggested as a potential explanation for employee investment allocation decisions in the context of 401(k) plans (Kamenica 2009). A second heuristic reflects the related possibility that participants selected goals based on a (relative) preference for competition. The possibility that variation in economic risk-taking might reflect differences in tastes for competition was advanced by Niederle and Vesterlund (2007). The table, however, provides no support for either of the heuristics.[20]

6.2. New Heuristic Explanation for Conservative Choice (Pairwise Heuristic)

What might explain conservative goal choice in the lab, the field, and potentially similar decision settings more broadly? We conclude by proposing a novel heuristic explanation for conservative menu-based financial decisions informed by exploratory pilot studies in which we asked participants to describe the details of their choice deliberations and our reading of the literature on decision-making and inference. The proposed heuristic broadly stipulates that a decision-maker selects an option from a menu through a succession of approximate, proximal, pairwise comparisons. Critically, the pairwise comparisons prompt systematic errors in relative inference leading the decision-maker to underestimate the likelihood of the riskier event and increase the likelihood of more conservative choice.

We outline the heuristic more formally by returning to our earlier theoretical framework in which goal choice is represented in the context of a two-option menu with binary lotteries varying across low and high risk. For simplicity, we focus on the decision of a risk-neutral employee. The heuristic specifies that such an employee will select the high goal, over the low-risk alternative, only if the expected gain from shifting to the high goal—i.e., the potential gain in reward from attaining the high compared to the low goal, weighted by the contingent likelihood of attaining the higher goal—exceeds its expected cost—i.e., the potential loss associated with not receiving the low-goal reward weighted by that likelihood. The employee will therefore select the high goal if the following pairwise comparison is satisfied:

$$\left(\hat{s}_{h|l} * \Delta x_{h,l}\right) + \nu > \left(\hat{s}_{\neg h|l} * x_l\right)$$

Here, $\hat{s}_{h|l}$ indicates the perceived belief of attaining the high-goal contingent on attaining the low goal, $\hat{s}_{\neg h|l}$ is the perceived belief of not attaining the high-goal contingent on attaining the low goal, $\Delta x_{h,l}$ is the gain in rewards from attaining the high, relative to the, low goal, x_l denotes the low-goal reward, and ν is an additive classical error parameter denoting an error of +/- ν dollars.

[20] We tested the ability-sorting heuristic by asking participants to assess their grid-solving ability relative to other participants, mapping relative assessments to predicted goal choice by menu position (e.g., high relative ability predicts Goal 3 choice, average relative ability implies Goal 2 choice, etc.), and then comparing actual and predicted choices. We used a similar procedure to test the relative taste-for-competition heuristic.

If employees had unbiased and otherwise well-calibrated beliefs about goal attainment, then the pairwise comparison simply restates the utility-maximizing proposition with the addition of the noise allowance. Critically, however, the heuristic dictates that employees systematically neglect contingent probabilities. Specifically, we posit that individuals underestimate contingent pairwise likelihoods leading employees to systematically underestimate the financial value of switching to the high goal. We represent the bias through a likelihood function, $\hat{s}_{h|l} = ks_{h|l} = k\frac{s_h}{s_l}$, where k $\in[s_l\ ,1)$, such that an employee partially to fully neglects to adjust the posterior by the marginal probability of low goal attainment. Induced by a focus on the relative comparison, this systematic bias in inference increases the likelihood of conservative goal choice. While the present framing involves a menu of two-options, we propose that decision-makers apply the heuristic to moderately larger menus through successive, proximal, pairwise comparisons beginning at the low-risk option. For example, in a menu with three ordered options, as in GQ, the pairwise heuristic would dictate an initial comparison between Goals 1 and 2, followed by a comparison of Goals 2 and 3 for anyone initially rejecting Goal 1.

Motivating Evidence from the Literature. While a dispositive account of the psychological micro-foundations of the proposed heuristic are beyond the scope of this paper, its three defining behavioral assumptions—i.e., the assertion of pairwise comparisons, the allowance for some computational error, and systematic bias in relative inference—are rooted in widely supported understanding of decision-making processes. For example, while the use of proximal, pairwise, comparisons to reduce the complexity of menu-based choice has not been directly asserted, to our knowledge, the propensity of individuals to engage in relative, as opposed to separate, evaluation has been established across a variety of decision processes and contexts through experimental and neuroscientific evidence and has been increasingly recognized by economists (e.g., Bushong, Rabin, and Schwartzstein 2021). Similarly, the allowance for approximation or noise in decision-processes is a key component of numerous decision-making frameworks (e.g., Camerer 1989; Harless and Camerer 1994; Hey and Orme 1994). and systematic errors in inference (e.g., Kahneman and Tversky 1973; Peterson and Miller 1965; Slovic and Lichtensetin 1971; Cornet et al. 2010; see Benjamin 2019).

6.3. Experimental Evidence for Pairwise Heuristic (Experiment B)

We assess the plausibility of the proposed heuristic through two strategies. First, we sought evidence from an experiment as to whether individuals conformed to the decision-making precepts underlying the pairwise heuristic—i.e., reliance on proximal pairwise comparisons and underestimation of pairwise contingent probabilities—whether any inferential bias helped to predict goal-choice, and whether a de-biased menu designed to defuse heuristic thinking led increased optimality of choice relative to

standard benchmarks. Second, we assessed whether the pairwise heuristic explains a greater share of goal choice than previously discussed benchmarks in the lab and in the field.

Overview. We administered the experiment in July 2022 on the Qualtrics platform to 893 employed US adults, aged 25 to 65, recruited from Amazon Mechanical Turk. After introducing participants to the real-life GQ paradigm, we randomized the 82 percent of participants who successfully completed multiple comprehension checks to one of two experimental arms. The first arm, designed to test whether participants used the pairwise heuristic in goal choice, asked participants to make a hypothetical goal-choice decision from a representative GQ menu, queried details of their decision-making process and elicited both contingent and non-contingent beliefs as to goal attainment. The exploratory pilot studies led us to believe that participants would engage clearly-worded vignettes in a manner similar to how employees engaged actual GQ decisions in the field. The second arm, designed to test whether defusing the heuristic through menu-design increased the optimality of choice, randomized participants to select a goal from one of three menus that, while featuring the same representative goals and rewards from the first arm, varied in their display of information about participant likelihoods of attaining each goal. Finally, all participants were asked about their about their contingent or non-contingent beliefs in an unrelated domain. The appendix displays key screenshots from the experiment.

Procedural Detail. After completing the module introducing GQ, the first arm of the study asked participants to select a goal from a representative GQ menu as if they were actually an employee participating in a real-life GQ program. The menu enumerated three sales goals (105 units, 110 units, 115 units) and associated rewards ($150, $450, $900) that resembled a prototypical menu engaged by employees in percent-denominated GQ programs (of note, average rewards in percent-denominated programs were somewhat higher than all GQ programs).[21] To infuse the hypothetical goal choice with greater realism, we provided participants a series of fictional sales figures for the prior 14 periods—the distribution was constructed to generate an implied likelihood of goal attainment similar to that observed in the field. After participants selected their goal, we asked them to introspect as to how they arrived at their goal choice. Specifically, we asked participants to indicate which, if any, pairwise comparisons they made between goals during their deliberation (e.g., "At some point, I directly compared Goals 1 and 2"). Participants then proceeded to a module that elicited non-contingent and contingent beliefs of goal attainment. Specifically, we asked participants to estimate, on a scale from 0 to 100 percent, their likelihood of goal 3 attainment given certain knowledge of attaining goal 2 (using a text-box). We then asked for the analogous likelihood of attaining goal 2 assuming attainment of goal 1. Given

[21] The menu was intended as a prototype for those used in percent-denominated GQ programs. To generate rewards, we applied the modal rewards ratio (1-3-6) to the median goal 1 reward ($150, after rounding). The goals reflect the mean/median/modal 5-10-15 percent increases relative to baseline across percent-denominated GQ programs, applied to a 100-unit baseline.

the hypothesized difficulty of clearly eliciting a contingent expectation we piloted several different communication strategies before arriving at the one we used throughout the experiment.[22] Participants were then asked to separately estimate the non-contingent likelihood of attaining each goal (we instituted validation rules and error messages to disallow internally inconsistent beliefs).[23]

As in the first arm, the second arm of the study asked participants who had just completed the GQ introductory module, to make a decision from a GQ menu with the same representative goals and rewards as if they were a real-life participant in the program. However, rather than providing a fictitious history of prior sales, we randomized participants to one of three variations of the menu that differentially encouraged non-contingent, unbiased contingent, or biased-contingent evaluation. Specifically, a first menu (non-contingent) communicated that a participant's likelihood of goal attainment was 83 (goal 1), 74 (goal 2), and 65 percent (goal 3) (e.g., "You have an 83 percent chance of achieving Goal 1"). These likelihoods reflected the average attainment statistics from the field. A second menu (unbiased contingent) displayed the same likelihood for goal 1, but then displayed accurate contingent likelihoods, $\hat{s}_{h|l} = s_{h|l}$, for goal 2 ("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"). Finally, a third menu (biased contingent) once again displayed the non-contingent likelihood for goal 1 but displayed contingent likelihoods for goals 2 and 3 reflecting presumed bias of the form, $\hat{s}_{h|l} = s_h$.

Participants across both arms concluded the study by submitting a non-contingent or contingent forecast in a domain unrelated to goal choice. Specifically, we randomized participants to either separately forecast the likelihood that tomorrow's high temperate where they lived would meet or exceed 70, 80, and 90 degrees Fahrenheit (we administered the study in the summer) or to forecast the likelihood that tomorrow's high temperature would meet or exceed 90 degrees, given certain knowledge it would meet or exceed 80 degrees. This final module was intended to provide a between-subject test of the hypothesized bias in inference.

Results. The experiment, in which baseline goal choice was (0.24, 0.45, 0.31), yielded several pieces of evidence implicating the likely use of the proposed heuristic, or some similar strategy, by a substantial share of participants. First, we document significant evidence for two of the seminal process assumptions underlying the pairwise heuristic. Specifically, 86 percent of participants reported using pairwise comparisons to arrive at their goal choice and 93 percent of such participants made at least one

[22] For example, to elicit contingent beliefs of weather we asked: "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?"

[23] We hypothesized that asking for contingent beliefs and non-contingent beliefs in succession might lead to more consistent responses than those yielded by a between-subject comparison. Pilot studies suggest that while both within- and between- subject comparisons yield substantial bias, within-subject estimates generate bias of smaller magnitude.

proximal comparison (58 percent relied on only on proximal comparisons). More critically, we documented evidence for substantial and pervasive underestimation of contingent likelihood pairs. For example, when asked to forecast the likelihood of a 90+ degree high-temperature given knowledge that the high would meet or exceed 80 degrees, participants underestimated the contingent likelihood by 38 percent relative to that implied by non-contingent estimates (0.45 relative to 0.72). Participants also exhibited bias in their within-subject estimates of attainment, underestimating the contingent likelihood of attaining goal 3 | goal 2 by 22 percent (0.59 relative to 0.76) and the likelihood of attaining goal 2 | goal 1 by 23 percent (0.64 relative to 0.82).

Second, we find that the magnitude of the bias in estimates of contingent likelihood strongly predict optimal goal choice, as determined by a risk-neutral, subjective expected-utility benchmark. For example, we estimate a simple additively linear model of optimal goal choice, g_c^*:

$$g_c^* = \alpha + \theta_1 \hat{s}_1 + \theta_2 \hat{s}_2 + \theta_3 \hat{s}_3 + \gamma_1 \lambda_{3,2} + \gamma_2 \lambda_{2,1} + \varepsilon$$

where $\hat{s}_k$ indicates the perceived likelihood of attaining goal, *k*, as indicated by non-contingent elicitations, and $\lambda_{k,k-1}$ denotes the magnitude of the bias in perceived contingent likelihood associated with goal *k* and goal *k*-1 (for example, $s_{3|2} - \hat{s}_{3|2}$), as estimated from the within-subject elicitations. The regression estimates suggest optimal choice, which for 91 percent of participants was goal 3, is substantially predicted by the perceived likelihood of goal 3 attainment (b = 1.00, p < 0.001) and negatively predicted by the magnitude of the relevant contingent bias (b = -0.81, p < 0.001).[24] Given the 0.37 overall share of optimal choice, the estimates imply that eliminating the bias, which averages 17 percentage points across the sample, would increase optimal choice by 37 percent (i.e., -0.17 x -0.81 / 0.37). While the simple specification we estimate is not intended to represent underlying decision processes, the considerable partial correlation between the contingent error and optimal choice is evident across a variety of non-parametric specifications.

Finally, the experiment documents a marked increase in optimal choice in response to a menu designed to defuse the pairwise heuristic relative to other meus. Specifically, when engaging the menu displaying non-contingent likelihoods informed by empirical averages from the field, 40.8 percent of participants selected the EV-maximizing high goal. However, engaging a menu displaying unbiased contingent probabilities, 60.5 percent selected the EV-maximizing goal, a 48 percent relative increase in optimal choice (p = 0.002). As evidence suggesting the accuracy of our specific assumptions about the magnitude and form of the bias, participants who engaged the menu displaying contingent likelihoods with full bias, $k = s_l$, made optimal choices at a rate, 39.4 percent, statistically indistinguishable from the

[24] We exclude six participants with a non-unique prediction from the benchmark model.

non-contingent menu (p = 0.82) and the menu with no information display from the first arm (p = 0.55). The comparisons not only document that de-biased menus increased optimal choice by roughly 50 percent but that such menus reduced the choice of goal 1, the lowest-EV option, by 27 percent (14.4 to 10.5).

6.4. Descriptive Accuracy of Pairwise Heuristic – Lab and Field

While the experiment offers evidence for the mechanistic assumptions underlying the pairwise heuristic, the partial correlation between the magnitude of the bias and optimal choice, and substantial improvement in choice efficiency from de-biased menus, perhaps the most informative test of the pairwise heuristic is the accuracy with which it explains goal choice in the field and lab, particularly relative to other benchmarks. To carry out the descriptive analysis, we must specify additional details of the heuristic and its application to a GQ menu. For example, unlike the lab where we directly observe each participant's contingent beliefs, in the field we must make assume a particular functional form for the inferential bias. Informed by the empirical relationship between contingent and implied contingent beliefs in the lab, particularly for relatively probably events, and response to the experimentally varying menus, we assume that employees in the field are subject to a bias of the form $\hat{s}_{h|l} = s_h$. We refer back to the lab to gauge how the parametric assumption of the bias might adversely affect estimates of descriptive accuracy. Next, we specify noise allowances of either \$0, +/- \$25, or +/- \$50, i.e., $\nu = \in \{0, [-25, 25], [-50, 50]\}$, a range arguably spans reasonable degrees of computational imprecision.[25] As discussed earlier, we also assume that decision-makers apply the heuristic to a GQ menu by first comparing Goals 1 and 2, and then, if they proceed to Goal 2, comparing Goals 2 and 3. Finally, to avoid the risk of inflating optimal choice shares due to benchmarks that predict non-unique first-best choices, we characterize choice for the restricted set of decisions for which the heuristic predicts a non-unique choice along with the unrestricted set of decisions, where we treat any prediction-consistent choice as optimal.

The estimates of descriptive accuracy, reported in Table 7, document the heightened explanatory power of the pairwise heuristic relative to the standard baseline. In the lab, the heuristic, inputted with personalized contingent beliefs, explains 0.55, 0.57, or 0.56 of choice, across noise allowances in the restricted sample of decisions for which the heuristic delivers a unique prediction. These rates of explanation rise in the unrestricted sample, a more demonstrable rise with a greater noise allowance. The explanatory power of the heuristic implies an increase, relative to the SEU benchmark, of 49 to 65 percent, comparable to the increase implied by experimental response to the de-biased and contingent-

[25] A noise of allowance of \$50 in the context of the representative menu from the experiment is equivalent to 20 percent of the average difference in expected value between goals 2 and 3 and 24 percent of the average difference between goals 1 and 2. The allowance is equivalent to even larger shares of the differences in expected value across all menus in the field. Interpreted in terms of wage-based time-use, a \$50 noise allowance is equivalent to roughly two hours of effort given earnings of \$25/hour.

display menus. Among the mechanistic elements of the pairwise heuristic, the table alludes to the primary importance of the inferential bias followed by the allowance of noise (the interaction of the two also predicts greater explanatory power). As indicated by the small difference in descriptive accuracy between the baseline and the pairwise heuristic with no noise and no inferential bias, the assumption of successive pairwise comparisons beginning at the low goal does not seem critical to the success of the heuristic.

In the field, the pairwise heuristic increases explanatory power relative to the baseline benchmark by 26 to 46 percent assuming a noise parameter of $25 across the restricted (0.50 to 0.63) and unrestricted samples (0.50 to 0.73). While explanatory power increases by up to 66 percent, relative to baseline, given a noise allowance of $50, we caution that a noise allowance of this size produces unique choice predictions for only 42 (parametric bias) to 61 (no bias) percent of the sample. Comparing the characterization of choice across personalized and parameterized formulations of the bias in the lab suggests that our field estimates of explanatory power might not look much different, or might even be higher, had we observed personalized bias. As in the field, the assumption of inferential bias is primarily responsible for increasing explanatory power though the degree of noise allowance is more important than it was in the lab (due to the lower average stakes in the field relative to the representative menu in the lab). Across the lab and field, the explanatory power of the pairwise heuristic exceeds prior benchmarks.

6.5. Explaining Residual Goal Choice – Local Pairwise Heuristic

We interpret evidence from the experiment and descriptive analyses as suggesting that a moderate to large share of employees use a decision-strategy resembling the pairwise heuristic to select their goal. While the difference in optimal choice in the field under the standard benchmarks and our heuristic offers a credible lower bound for the share of employees using the heuristic as between 13 and 33 percent (i.e., the difference between 50 percent and 63 to 83 percent in Table 7), the experiments imply a potentially far higher upper bound, as many choices seemingly consistent with a benchmark may have nevertheless been generated heuristically. For example, in the lab, the pairwise heuristic with $v = 25$ explained 92 percent of unrestricted choices also explained by the risk-neutral SEU benchmark and 30 percent of choices unexplained by the benchmark. Inversely, the benchmark explained 69 percent of choices also explained by the heuristic and only 8 percent of choices unexplained by the heuristic.

Notably, a significant fraction of individuals doesn't seem to adhere to the standard benchmark, or the alternatives we tested, inclusive of the proposed heuristic. While some of these decisions, in the lab and the field, likely reflect confusion, inattention, or an otherwise random choice process, we can make informed speculations as to the decision strategies that may have produced residual choice. First, we suspect that some individuals may have applied the pairwise heuristic to a narrowly focused portion of the menu, that is, after ruling out one of the options, most likely the high goal. Consider that approximately

one-quarter of experimental participants who made choices unexplained by the benchmark or heuristic reported only comparing goals 1 and 2—an indication fully consistent with final goal choice. Applying the pairwise heuristic only to goals 1 and 2 explains all but one of these decisions. While we do not know why an individual might initially rule out the highest goal, narrowly engaging a menu seems reasonable in the context of larger ordered-menus where pairwise comparisons across the entire menu might be overly effortful or engage undesired choices (e.g., an insurance consumer who knows they do not want a high-coverage plan). Finally, based on the first experiment where we observe multiple choices per individual, we surmise that a modest share of individuals was intent on selecting the lowest goal regardless of the goal's relative economic value. This could reflect a variety of decision motives such as a desire to maximize the likelihood of gain, without regard to its magnitude, or the desire to minimize risk.

7 IMPLICATIONS FOR POLICY AND PROGRAMS

While the simplicity of GQ positions it favorably as a setting for research, we speculate that individuals apply heuristics, perhaps resembling the pairwise heuristic described in this paper, to other menu-based decisions involving risky choice such as portfolio allocation or insurance choice. One distinguishing feature of many menus of economic interest is the presence of non-contingent costs. That is, in contrast to GQ where a decision-maker comparing Goals 2 and 3 can freely ignore the potential failure to attain Goal 2 (since such a failure yields no reward regardless of goal choice) in settings such as insurance, plans differ in cost that does not depend on the realization of risk (e.g., insurance premium). The neglect of contingent likelihood, associated with our heuristic, predicts that when facing menus with non-contingent costs, a decision-maker will underestimate the likelihood of a large, relative to a small, risk in the context of pairwise comparisons but also predicts insufficient adjustment by baseline risk.

As illustration, consider a highly stylized setting in which an individual must purchase insurance for a new home (e.g., due to a lender mandate). Suppose that practically she must choose from a menu of two policies (Low Coverage and High Coverage) that are identical but for cost-sharing (Low: $1,000 deductible; High: $500 deductible) and annual cost (Low: $600 premium; High: $700 premium). We further assume a 4 percent chance of (covered) damage to the home, that any damage is in excess of $1,000, and the consumer is risk-neutral. The pairwise heuristic, in this scenario, stipulates that the consumer would select a plan by assessing, with some noise allowance, *v*, whether the potential benefit of shifting from low to high coverage, $\Delta x_{low,high}$, in the event of a claim, exceeds the difference in policy costs, $\Delta p_{low,high}$. The consumer would therefore select the high coverage plan if:

$$\hat{s}_{claim} * E\left(\Delta x_{low,high} \mid claim\right) + \nu > \Delta p_{low,high}$$

However, as earlier, the heuristic implies the (partial) neglect of the contingency, such that $\hat{s}_{claim} = \theta \hat{s}_{claim}$ where, $\theta \in (0,1)$. In this simplified setting, the contingency neglect takes the form of base-rate neglect—an inferential bias for which there is considerable theoretical and empirical evidence (see Benjamin 2019). If one were to complicate the cost distribution of potential claims by specifying the possibility of an either small or large claim, for example, then the pairwise comparison, predicted by the heuristic could generate biases in multiple relative inferences, leading to a pattern of sorting into an ordered-plan menu not predicted by underlying risk or risk perceptions.

To explore whether heuristic choice involving contingency neglect can help to explain insurance choice, we administered a final online experiment to test whether menus that either encouraged or defused contingent inference shifted hypothetical insurance choice. We asked 435 US adults, aged 25 to 55 years, recruited from Amazon Mechanical Turk to imagine they had to select an insurance plan for their new home (we indicated insurance was lender-mandated) from a menu of three options, adapted from an example in Sydnor (2010), varying only in their deductible and annual cost (with full coverage beyond the deductible): (1) Basic Plan: $1,000 deductible/$616 premium, (2) Medium Plan: $500 deductible/$716 premium, (3) Premium Plan: $250 deductible/$803 premium).

We randomize participants to render their choice from one of 4 versions of the menu. While each featured the same plans, the menus varied in the display of loss probabilities in a manner intended to either encourage or discourage contingent/heuristic thought. Specifically, assume a 4 percent overall likelihood of damage (the average claim rate reported in Sydnor, 2010), a 3 percent likelihood of severe damage (more than $2500), and a 1 percent likelihood of non-severe damage (less than $2500). Given this assumed distribution of potential loss, the four menus included: a (1) Baseline menu with no loss information, a (2) Contingent Loss + No Base Rate displaying the 75 percent conditional likelihood of a severe loss, a (3) Contingent Loss + Base Rate displaying the 4 percent loss likelihood and the 75 percent conditional likelihood of a severe loss, and a (4) Non-Contingent menu displaying the 96 percent likelihood of no loss. The second menu was intended to most strongly encourage contingent thought (by displaying the contingent likelihood and no base rate) while the final menu was intended to most strongly encourage non-contingent thought (by displaying only the base rate in terms of no loss).

While the menus vary in both the framing of information and its presence, without sharply inflated perceptions of loss, relative to the empirical average, it is difficult for the expected-utility framework to rationalize any choice other than the cheapest plan (see discussion in Sydnor 2010). For example, the most pessimistic, relative to the most optimistic, assumption about the cost of non-severe loss (e.g., $2,499 versus $1) given the indicated cost of severe loss, would only reduce the difference in expected value between the Basic and Medium plans by roughly $5, implying the informational near-equivalence of the final two menus. As reported in Appendix Table A3, the experiment revealed that

participants chose plans across the menu, despite the low expected economic value of the medium and premium plans. When faced with a menu encouraging contingent engagement (2 and 3), participants were significantly less likely to select the cheaper, high-deductible plan than menus discouraging contingent engagement (4). This experiment, in conjunction with the evidence from GQ, suggests how substantially conservative, and heterogeneous, choice could arise even in the presence of plausible risk preferences and well-calibrated beliefs.

8 CONCLUSION

We describe new evidence on the magnitude of financial risk-taking and its underlying motives. Our evidence describes the decisions of several thousand employees in the context of a popular employee reward program. We see this setting as uniquely helpful for understanding risky choice given the diversity of the decision makers, the wide-ranging financial magnitudes, the simplicity of the choice environment, and our visibility into contemporaneous employee beliefs. A central finding is to document substantial risk aversion, and heterogeneity, in the goal choices of employees, resulting in an average unrealized gain equivalent to 30 percent of potential rewards. The excess conservativism of employees was robust to reward size and employee tenure. We proceed to show that conservative goal choice cannot be explained by utility-based preferences for risk or common behavioral departures from the expected utility framework (e.g., biased beliefs, non-linear decision weights, or gain-loss utility). Across models we tested, none explained substantially more than one-half of employee choices. An online goal-reward paradigm online, in the context of an incentive-compatible effort task, replicated the conservative choice from the field in a setting with dollar-denominated rewards, verified comprehension of the paradigm, and little scope for signaling, reputational concerns, or high cost of effort. The experiment corroborated the challenges of explaining choice from the benchmark models assessed in the field.

We proposed a novel heuristic explanation for the observed pattern of choice based on pilot studies exploring the phenomenology of choice. The pairwise heuristic stipulates that individuals engage the choice menu through a series of proximal, approximate, pairwise comparisons. The heuristic posits that comparisons are subject to inferential error, however, which leads individuals to effectively underestimate the likelihood of attaining the higher goal. It is this underestimation that helps to explain the diversity and conservatism of choice. We administered a second experiment that corroborated the mechanistic presumptions of the heuristic, revealed a correlation between the inferential bias and optimal choice (controlling for other beliefs), and documented the substantial responsiveness of participants to menus designed to de-bias the heuristic by discouraging contingent inference. Perhaps the most compelling evidence for the heuristic is that it explains a greater share of choice in the lab and in the field than any of the prior benchmarks.

Beyond providing new evidence on the prevalence and motives of financial risk-aversion, a practical implication of this research is that the possibility that household financial decisions in other menu-based domains, such as insurance or investment decisions, may be, at least in part, generated from the application of heuristics rather than utility-based risk preferences. As an illustrative example, we administered a final online experiment to test how hypothetical home insurance choice from a menu encouraging non-contingent thought differed from a menu encouraging contingent thought. The responsiveness of choice to varying menu designs supports the possibility that the substantial risk aversion and high diversity that characterizes insurance choice in many settings may reflect the variable application of menu-based heuristics rather than heterogeneity in risk, risk preferences, or wealth. We hope that further work will clarify the specific cognitive processes that underlie the errors in contingent inference we document and will further explore the extent to which contingency-based heuristics might explain risky choice in other consequential settings.

8 REFERENCES

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

Barseghyan, L., F. Molinari, T. O'Donoghue, and 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, and J. C. Teitelbaum (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, 56(2), 501–64.

Barseghyan, L., J. Prince, and J. C. Teitelbaum (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*, 101(2), 591–631.

Beetsma, R. M. W. J., & Schotman, P. C. (2001). Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo. *The Economic Journal*, 111(474), 821–848.

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

Bombardini, M., and Trebbi, F. (2012). Risk aversion and Expected Utility Theory: An Experiment with Large and Small Stakes. *Journal of the European Economic Association*, 10(6), 1348–1399.

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

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

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, *2*(1), 61-104.

Damasio, A. (2005). Descartes' Error: Emotion, Reason, and the Human Brain. New York, NY: Penguin.

Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica*, 59(3), 667–686.

Handel, B. R. and J. Kolstad (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8), 2449–2500.

Harbaugh, W. T., K. Krause, and L. Vesterlund (2010). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 120(545), 595–611.

Heath, C., Larrick, R. P., and G. Wu. (1999). Goals as Reference Points. *Cognitive Psychology*, 38(1), 79–109.

Kahneman, D., and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.

Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1), 119-131.

Koszegi, Botond, and Matthew Rabin. (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, 121 (4): 1133–65.

Koszegi, Botond, and Matthew Rabin. (2007). "Reference-Dependent Risk Attitudes." *American Economic Review*, 97 (4): 1047–73.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90, 426–432.

Loewenstein, G., and Lerner, J. (2003). The role of affect in decision making, in Handbook of Affective Science, eds R. Davidson, K. Sherer, and H. Goldsmith (Oxford: Oxford University Press), 619–642.

Loomes, G., and Sugden, R. (1982). Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal*, 92, 805–824.

Niederle, Muriel and Lise Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much?, *The Quarterly Journal of Economics*, 122(3): 1067-1101.

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

Pratt, John W. (1964). "Risk Aversion in the Small and In the Large." *Econometrica*, 32 (1): 122–36.

Prelec, Drazen. (1998). "The Probability Weighting Function." *Econometrica*, 66 (3): 497–527.

Starmer, Chris. 2000. Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature*, 38 (2): 332–82.

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.

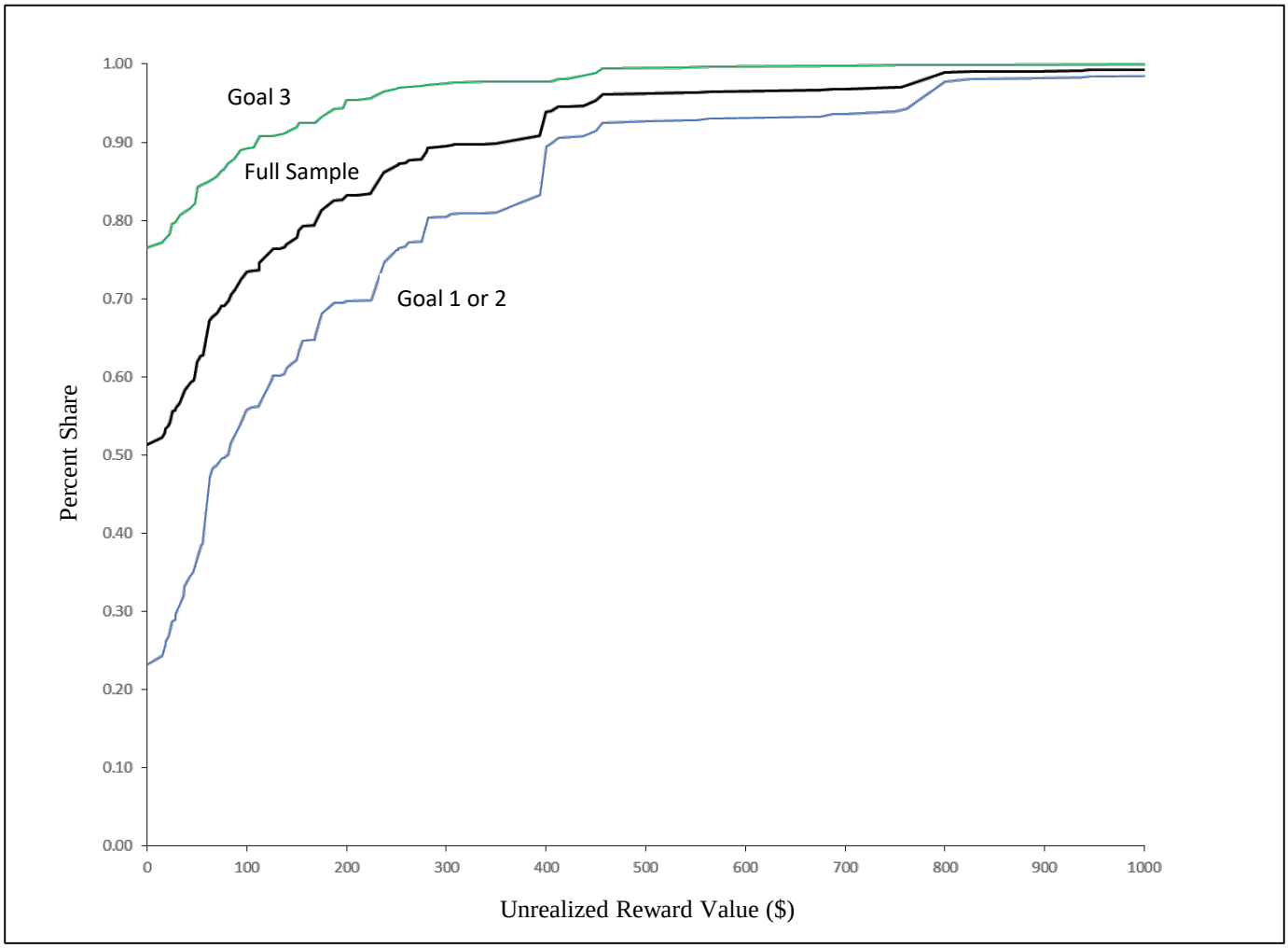
Sydnor, Justin. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics*, 2 (4): 177–99.

Tversky, Amos, and 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), 547–567.

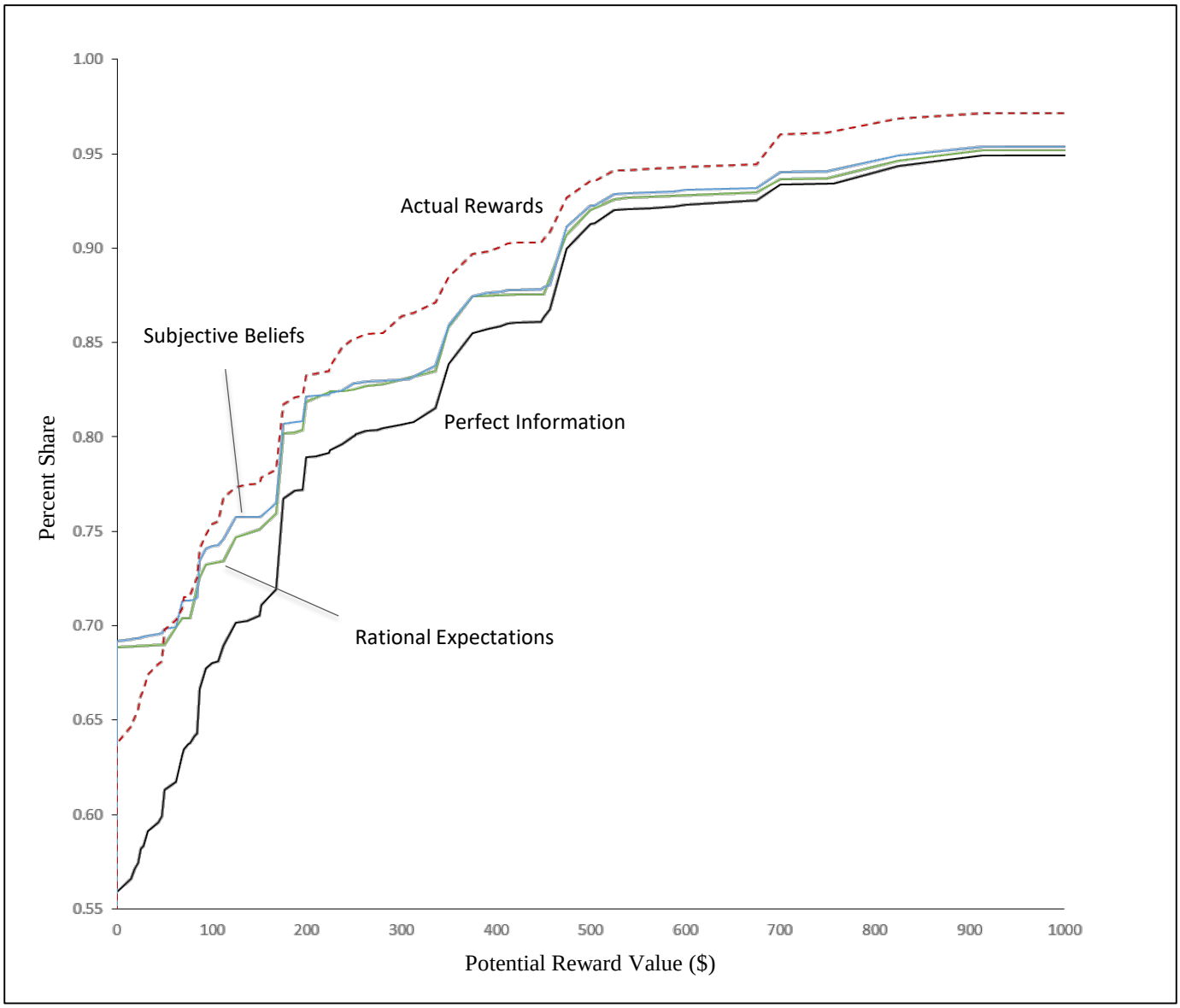
von Neumann, J., & O. Morgenstern. (1947). Theory of games and economic behavior (2nd ed.). Princeton, NJ: Princeton University Press.

Figure 1.
Cumulative Distribution of Unrealized Rewards relative to Ex Post Optimal Choice



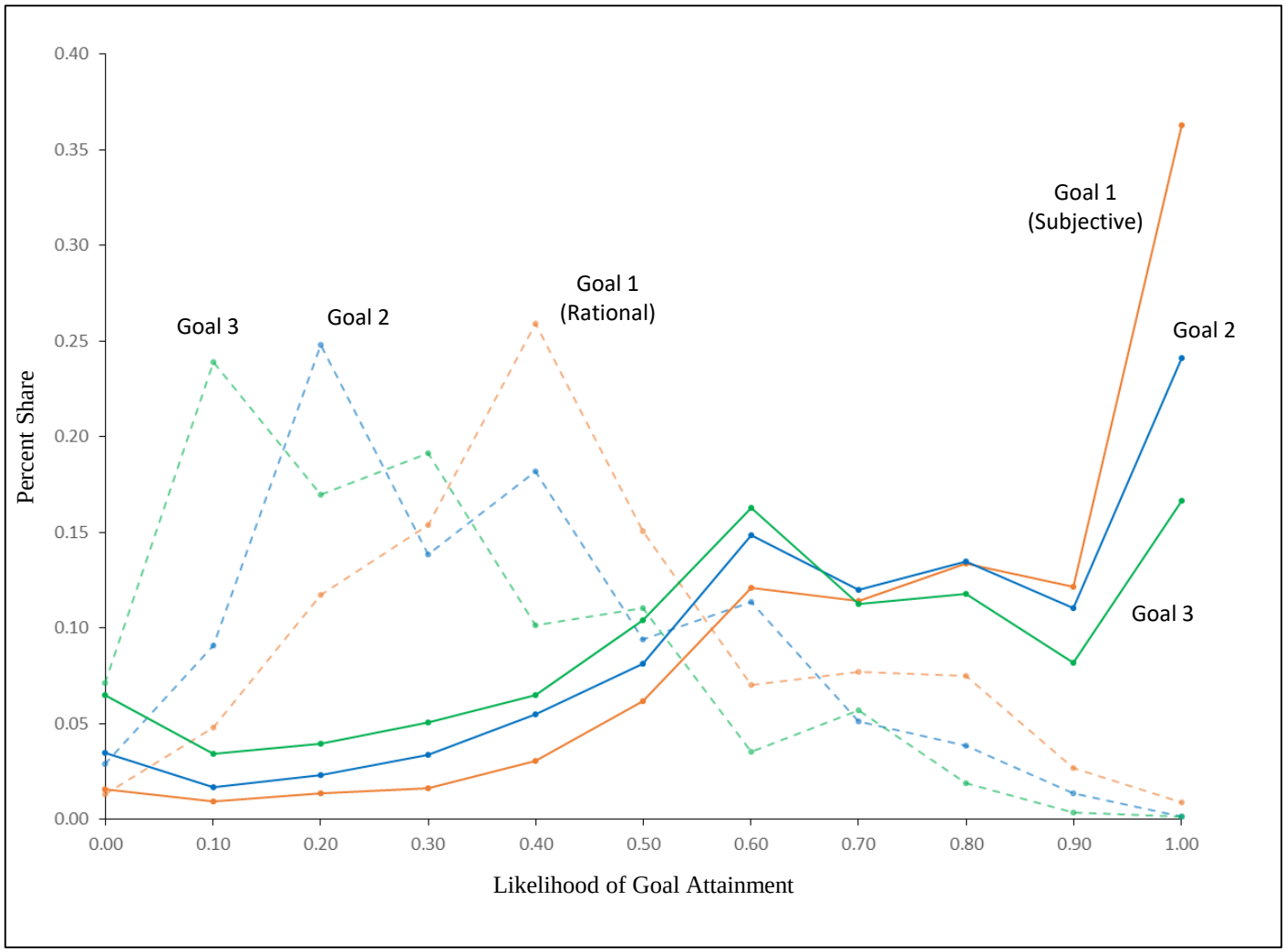
Notes: This figure depicts the cumulative distribution of unrealized rewards overall and separately by goal choice for the 8,800 employees whose productivity met or exceeded the Goal 1 threshold. An unrealized reward refers to the difference between an employee's actual earned reward and the counterfactual reward an employee could have earned if they had chosen ex post optimally. By definition unrealized rewards cannot be negative. While the figure censors unrealized rewards at $1,000 for clarity, a small share of employees had unrealized rewards in excess of $1,000, with a maximum unrealized reward of approximately $2,800.

Figure 2.
Cumulative Distribution of Actual and Counterfactual Rewards under Risk Neutral Expected Utility Benchmark by Information Regime



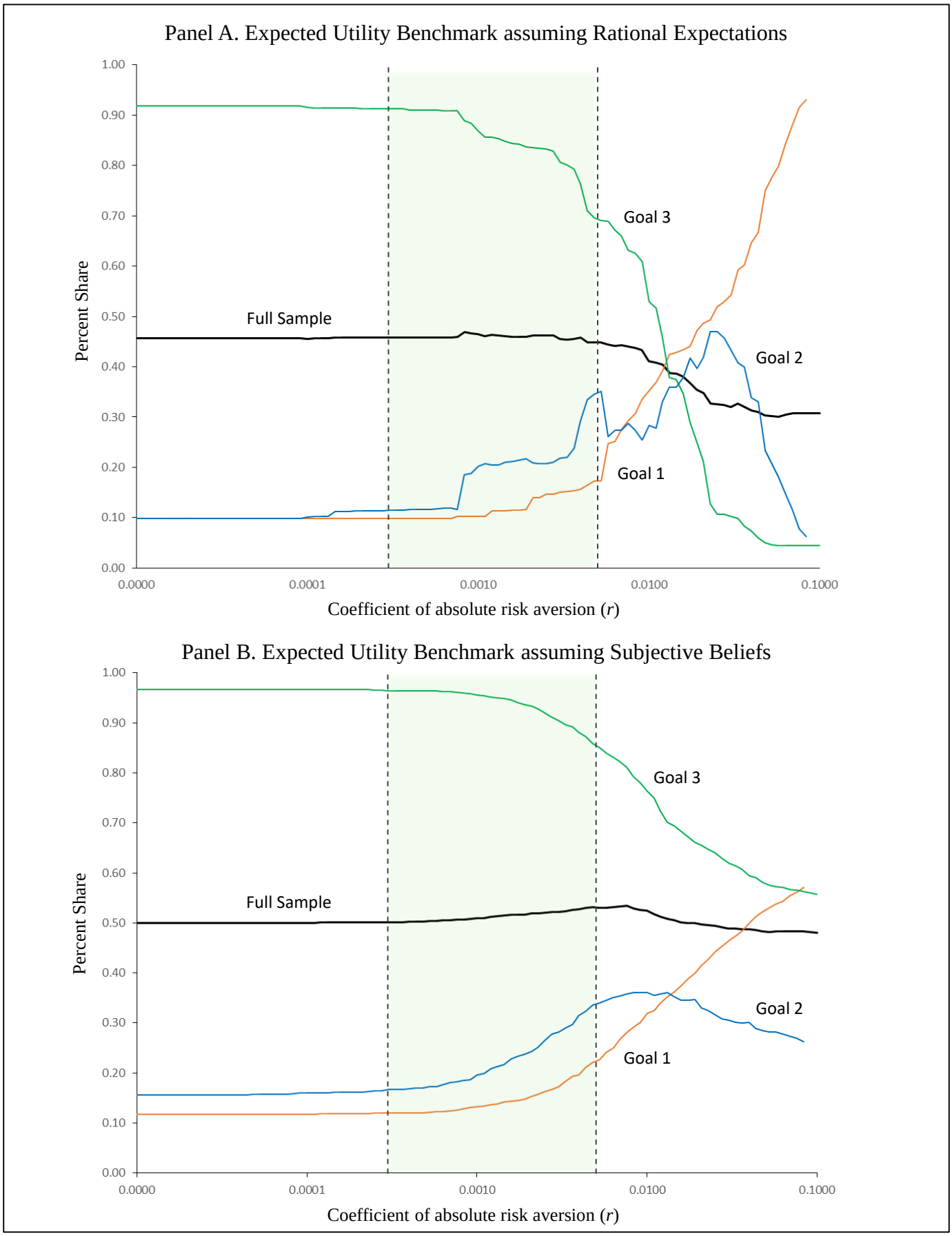
Notes: This figure depicts the cumulative distribution of actual rewards and counterfactual rewards under expected utility for risk neutral employees by information regime. Specifically, the dashed red line indicates the distribution of actual earned rewards, the black line indicates the distribution 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.

Figure 3.
Distribution of Rational Expectations and Subjective Beliefs of Goal Attainment



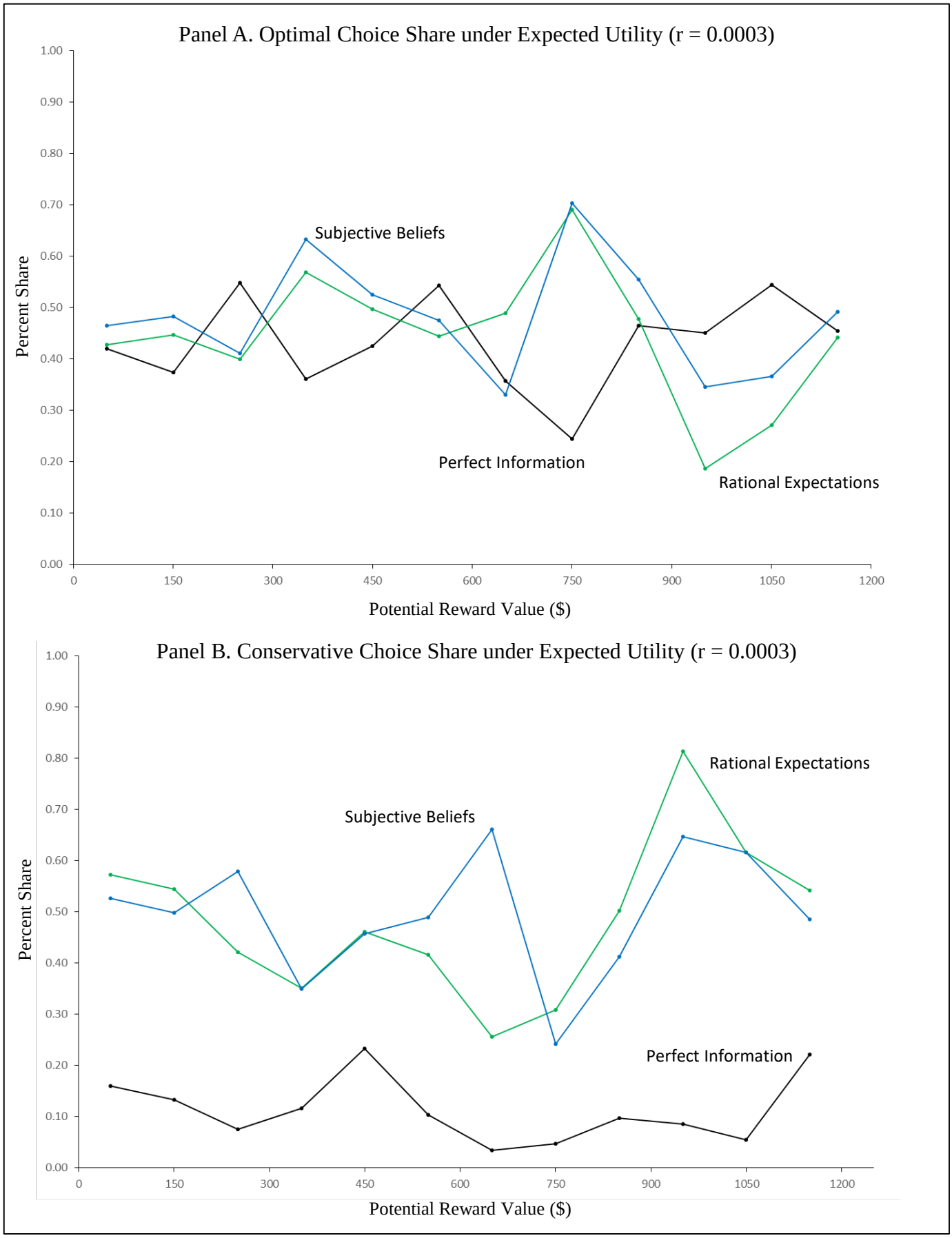
Notes: This figure compares the distributions of rational expectations and subjective beliefs of goal attainment for each goal. We assign employee- and goal-specific rational expectations by adjusting the ex post average rate of goal attainment at the group-level by employee age and gender, as estimated 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.

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



Notes: This figure depicts the share of optimal choice overall and separately by goal choice under expected utility across varying levels of the CARA risk aversion parameter, r, and information regimes. Specifically, 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 characterization of choice under the assumption of subjective beliefs. The shaded region denotes the range of substantial but still plausible risk aversion, $r \in [0.0003, 0.05]$.

Figure 5.
Optimal and Conservative Choice under Expected Utility Benchmark by Potential Reward and Information Regime



Notes: This figure reports the share of optimal (Panel A) and conservative (Panel B) choice by potential reward value under expected utility with moderate risk aversion (r = 0.0003) for three different information regimes (perfect information, rational expectations, and subjective beliefs). For each employee, the potential reward is defined as the highest possible reward they can earn in the program, or the reward associated with Goal 3. The plots group data into $50-bins of potential reward and, for clarity, censors potential rewards at $1,150.

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)	308	-	-
	(277.3)		
Employees per Program (Average)	1158	-	-
	(947.7)		
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	466.5	150.1	746
	(481.5)	(57.8)	(516.8)
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. Specifically, Panel A describes the number and size of programs across the overall analytic sample, while Panel B describes average program duration and average potential reward values at the employee-level. Potential reward value refers to the largest reward an employee can potentially earn in the program, or alternatively, the value of the Goal 3 reward. Panel C summarizes demographic details of employees including age, gender, tenure, and approximate salary for all employees 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 inference from first name, and approximate salary using program-level averages for those programs for which data was available.

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

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

Notes: This table summarizes goal choice, productivity, and goal attainment for the overall sample and separately by employee goal choice. Specifically, Panel A summarizes goal choice and average potential rewards, Panel B summarizes employee productivity relative to baseline and to Goal 3, and Panel C summarizes goal attainment and average earned rewards. Potential reward value refers to the largest reward an employee could potentially earn in the program, or alternatively, the value of the Goal 3 reward. The summary of productivity relative to baseline excludes the 18 percent of employees without baseline data.

Table 3.
Employee Beliefs and Confidence of Goal Attainment

		By Goal Choice		
	All	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 with respect to goal attainment for the overall sample and separately by employee goal choice. Specifically, Panel A successively summarizes beliefs of goal attainment under rational expectations and then under subjective beliefs. 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 tractability, we adjust any belief of 0 or 100 percent to 1 and 99 percent, respectively. Panel B summarizes employee under/over confidence as conveyed by the average ratio of subjective beliefs and rational expectations, such that a ratio > 1 indicates overconfidence. To minimize the effects of outliers, we Winsorized ratios by capping outliers below the 5th percentile and above the 95th percentile. Finally, the panel reports relative under/over confidence across specific goal pairs, as conveyed by the average ratio of Winsorized under/over confidence.

Table 4.
Goal Choice Characterization for Expected Utility Benchmarks

				Absolute Risk Aversion (CARA)					
	Risk Neutral			Rational Expectations			Subjective Beliefs		
	Perfect	Rational	Subjective	r = 0.0003	r = 0.005	r [0, 0.005]	r= 0.0003	r = 0.005	r [0, 0.005]
Panel A. Characterizing Goal Choice									
Optimal Choice	0.51	0.45	0.50	0.45	0.44	0.55	0.50	0.53	0.59
Conservative Choice	0.31	0.49	0.48	0.49	0.38	--	0.48	0.39	--
Aggressive Choice	0.18	0.06	0.02	0.06	0.17	--	0.02	0.08	--
Expected Reward | Chosen Goal	274	109	214	109	109	--	214	214	--
Expected Reward | Optimal Choice	381	146	275	146	124	--	275	247	--
Maximum Expected Loss	2813	1435	2272	1435	998	--	2272	2272	--
Average Potential Reward	466	466	466	466	466	--	466	466	--
Average Unrealized Reward | Conservative Choice	272	139	120	139	122	--	118	104	--
Panel B. Optimal Choice Share by Reward and Tenure		0.30							
Potential Reward Value									
Highest Quartile	0.53	0.42	0.48	0.42	0.39	0.62	0.49	0.55	0.69
Lowest Quartile	0.48	0.44	0.48	0.44	0.44	0.46	0.48	0.48	0.49
Employee Tenure									
Highest Category [10+ Years]	0.49	0.39	0.45	0.40	0.40	0.51	0.46	0.53	0.60
Lowest Category [< 1 Year]	0.52	0.44	0.47	0.44	0.44	0.58	0.47	0.50	0.62

Notes: This table characterizes the efficiency of employee goal choice under expected utility across a range of assumptions regarding CARA risk preferences and employee beliefs (perfect, rational, and subjective). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model and additionally reports average expected and unrealized rewards conditional on choice. Panel B reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience. Characterization of choice under the perfect information benchmark excludes employees who did not attain Goal 1 (characterization under other benchmark models rely on the entire sample). The blank cells reflect the inability to uniquely characterize aggressive and conservative choices for benchmarks involving flexible values of r .

Table 5.
Goal Choice Characterization for Non-Standard Benchmarks

				Gain-Loss Utility [$\alpha = 0.88$, $\lambda = 2.25$]		
	Baseline Subjective EU [r = 0.0003]	Noisy Belief Subjective EU [20% Error Band]	Rank-Dependent Subjective EU [Prelec $\alpha = \beta = 0.65$]	Gain-Loss Only [RP = g+1; $\eta = 0$]	Composite Utility [RP = Chosen g; $\eta = 1$]	Counterfactual Regret [RP = Counterfactual; $\eta = 3$]
Panel A. Characterizing Goal Choice						
Optimal Choice	0.50	0.54	0.47	0.55	0.59	0.50
Conservative Choice	0.48	0.45	0.52	0.23	0.24	0.48
Aggressive Choice	0.02	0.01	0.01	0.22	0.17	0.02
Expected Reward | Chosen Goal	214	229	188	214	214	214
Expected Reward | Optimal Choice	275	273	244	230	244	275
Maximum Expected Loss	2272	2029	2135	2272	2272	2272
Average Potential Reward	466	466	466	466	466	466
Average Unrealized Reward | Conservative Choice	118	110	120	126	129	120
Panel B. Optimal Choice Share by Reward and Tenure						
Potential Reward Value						
Highest Quartile	0.49	0.55	0.44	0.54	0.61	0.48
Lowest Quartile	0.48	0.51	0.46	0.53	0.55	0.48
Employee Tenure						
Highest Category [10+ Years]	0.46	0.52	0.42	0.53	0.59	0.46
Lowest Category [< 1 Year]	0.47	0.51	0.44	0.55	0.59	0.47

Notes: This table characterizes the efficiency of employee goal choice under a range of non-standard benchmark models. The first column of the table reproduces a baseline characterization from the expected utility benchmark assuming moderate CARA risk aversion (r = 0.0003) and subjective beliefs. The second column characterizes choice assuming a rank-dependent utility function, which modifies the baseline benchmark by applying nonlinear weights to potential outcomes using the Prelec function (1998). The final three columns characterize choice under exemplar benchmarks with gain-loss utility (but not nonlinear weighting) using the Kahneman and Tversky (1979) value function (see text for details). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model and additionally reports average expected and unrealized rewards conditional on choice. Panel B reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience. Characterization of choice under the perfect information benchmark excludes employees who did not attain Goal 1 (characterization under other benchmark models rely on the entire sample).

Table 6.
Goal Choice Characterization for Standard and Non-Standard Benchmarks ─ Experimental Paradigm (Study A)

						Alternative Heuristic Choice Models	
	Risk-Neutral Expected Utility [RE, r = 0]	Subjective Expected Utility [r = 0.0003]	Rank-Dependent Subjective EU [Prelec $\alpha = \beta = 0.5$]	Consumption + Gain-Loss Utility [RP = g+1; $\eta = 3$]	Noisy Belief Subjective EU [20% Error Band]	Contextual Sorting	Taste for Competition
All Menus (6/6)	0.13	0.04	0.03	0.18	0.24	0.10	0.09
Nearly All Menus (5+/6)	0.29	0.16	0.13	0.40	0.42	0.12	0.10
All 3 Goal Menus (4/4)	0.15	0.16	0.09	0.25	0.29	0.13	0.12
All 4 Goal Menus (2/2)	0.35	0.10	0.10	0.40	0.44	0.15	0.15

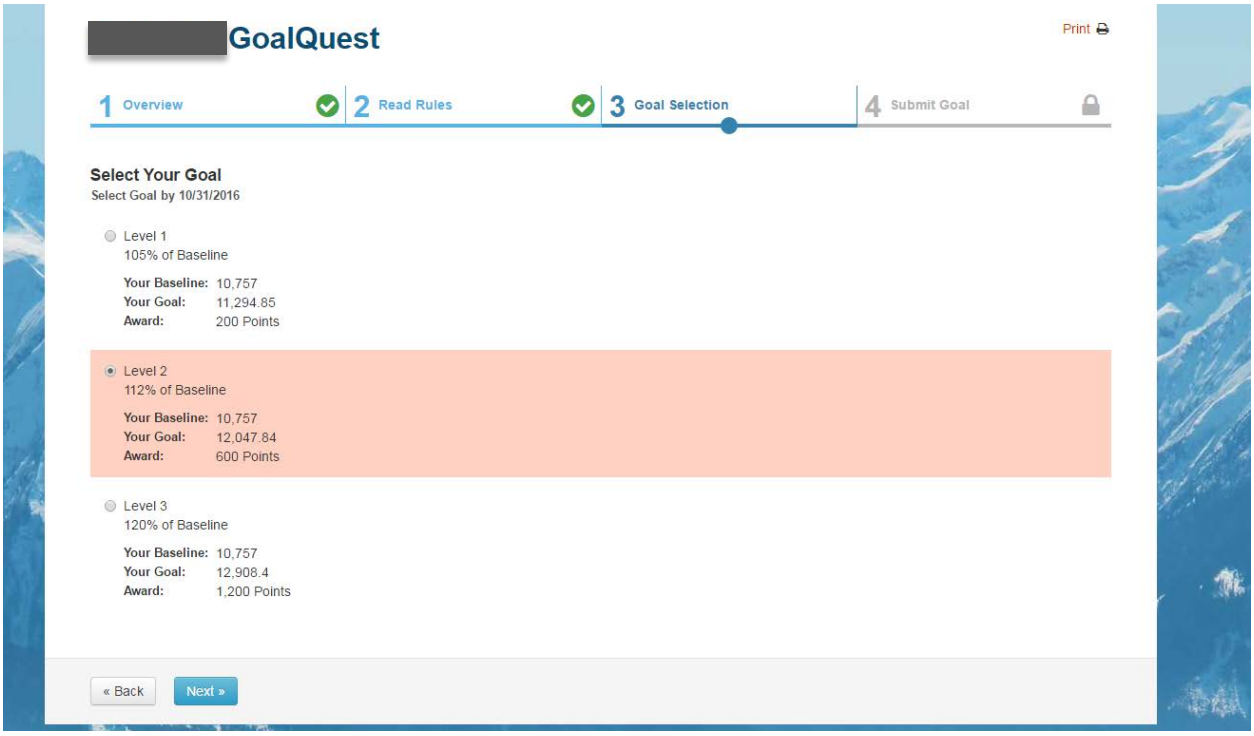
Notes: This table characterizes the share of optimal (Panel A) and conservative (Panel B) goal choice for experimental participants under a range of standard and non-standard benchmark models. The experiments asked each participant to indicate their goal-choice across six distinct menus in the context of an online effort task. The first column characterizes choice for a baseline model of risk-neutral expected utility assuming rational expectations. The remaining columns all assume subjective beliefs and characterize choice for, from left to right, a subjective expected utility model assuming moderate risk aversion (r = 0.0003), a fuzzy expected utility model that permits 10 percent +/- errors in calculations of expected utility (r = 0.0003), a rank-dependent model of expected utility that applies nonlinear weights to potential outcomes using the Prelec function (1998), and a model of composite gain-loss utility with the goal + 1 reward as the reference point and n = 3. A final set of columns characterizes choice for a set of heuristic-choice models whose details we describe in the text. The blank cells in Panel B reflect the inability to uniquely characterize conservative choices for the benchmark involving fuzzy expected utility.

Table 7.
Characterizing Accuracy of Pairwise Heuristic in the Lab and Field

		Pairwise Heuristic - Bias in Contigent Inference								
	Baseline	None			Personalized			Parameterized		
Decision Sample	SEU [r=0]	[v = $0]	[v = $25]	[v = $50]	[v = $0]	[v = $25]	[v = $50]	[v = $0]	[v = $25]	[v = $50]
Field Data										
Decisions with unique first-best options	0.50	0.48	0.50	0.53	--	--	--	0.57	0.63	0.67
All decisions	0.50	0.51	0.58	0.63	--	--	--	0.57	0.73	0.83
Experiment B										
Decisions with unique first-best options	0.37	0.36	0.37	0.38	0.55	0.57	0.56	0.53	0.54	0.57
All decisions	0.37	0.39	0.39	0.41	0.57	0.60	0.61	0.53	0.57	0.62

Notes: This table characterizes the share of optimal goal choice in the field and Experiment B under the pairwise heuristic across varying formulations of the inferential bias and noise allowance. The table reports optimal choice shares for both a restricted sample of decisions with a unique first-best goal choice and an unrestricted sample of all decisions.

Appendix Figure A1.
Sample Image of GQ Goal Selection Interface



Appendix Table A1.
Descriptive Accuracy of Gain-Loss Utility Models by Candidate Reference Point

	Gain-Loss Utility ($\alpha = 0.88$)			Consumpton + Gain-Loss Utility ($\alpha = 0.88$, $\lambda = 2.25$)			
Candidate Reference Points	$\lambda = 1.50$	$\lambda = 2.25$	$\lambda = 3.00$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 5$
Panel A. Prospect Independent							
Status Quo (0)	0.50	0.50	0.50	0.50	0.50	0.50	0.50
High Probability (Goal 1)	0.52	0.54	0.55	0.51	0.51	0.50	0.50
Compromise Goal (Goal 2)	0.50	0.52	0.52	0.50	0.50	0.50	0.50
Maximum Reward (Goal 3)	0.49	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	0.50
Panel B. Prospect-Dependent							
Reward of Chosen Goal	0.29	0.29	0.29	0.59	0.57	0.56	0.54
Expected Value of Chosen Goal	0.40	0.26	0.26	0.54	0.51	0.50	0.50
Reward of Chosen Goal + 1	0.55	0.55	0.55	0.53	0.53	0.53	0.52
Reward of Chosen Goal - 1	0.46	0.43	0.42	0.58	0.56	0.54	0.53
Regret (Expected Max Counterfactual)	0.50	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 A2.
Summary of Goal Choice, Beliefs, and Attainment — Experimental A (3-choice Menus)

		Sample Restricted by Goal Choice		
	All	Goal 1	Goal 2	Goal 3
Panel A. Goal Choice				
Goal 1	0.43	1.00	0.00	0.00
Goal 2	0.32	0.00	1.00	0.00
Goal 3	0.25	0.00	0.00	1.00
Number of Subjects	277	207	201	123
Number of Choices (3-choice menus)	1108	471	356	281
Panel B. Employee Beliefs				
Expected Performance	11.0	8.5	11.3	14.8
Expected / Actual Performance	1.5	1.4	1.4	1.6
Panel C. Goal Attainment				
Goal 1	0.68	0.52	0.78	0.84
Goal 2	0.51	0.28	0.60	0.76
Goal 3	0.27	0.08	0.23	0.64
Earned Reward (Average)	0.12	0.05	0.12	0.22
Earned Reward (Average) | Goal Attainment	0.20	0.09	0.20	0.35

Notes: This table summarizes goal choice, beliefs, and goal attainment overall and separately by goal choice for participants of the goal-choice experiments. The experiments asked each participant to indicate their goal-choice across six distinct menus in the context of an online effort task. Specifically, Panel A summarizes goal choice, Panel B summarizes actual and expected performance, and Panel C summarizes goal attainment and earned rewards.

Appendix Table A3.
Home Insurance Policy Choice across Display of Contingent Likelihood

	Menu Display			
Plan Type	Baseline	Contingent Loss (No Base Rate)	Contingent Loss (Base-Rate)	Non-Contingent
Basic Plan [D: $1,000, P: $616]	0.40	0.23	0.35	0.54
Medium Plan [D: $500, P: $716]	0.41	0.50	0.39	0.38
Premium Plan [D: $250, P: $803]	0.19	0.26	0.26	0.08
Expected Total Plan Cost	717	729	726	696

Notes: This table reports the average plan choice shares for participants from Experiment C (N = 435). The deductible and premium for each plan is diplayed in brackets. The expected total plan cost refers to the total out-of-pocket cost assuming a 3 percent chance of a loss exceeding $2,500 and a 1 percent chance of a loss of $500.