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: February 2023

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 rare 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. We estimate nearly one-half of employees selected a goal lower than the EV-maximizing benchmark, assuming rational expectations, resulting in a 46 percent average loss of potential rewards. This conservative goal choice persisted across diverse financial stakes ($69 to $4,500) and employee experience. We additionally show that conservative choice cannot be explained by a standard expected utility (EU) model with plausible risk preferences or through common departures from EU 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, corroborate limits of EU-based explanations, and rule out potential confounds through an incentive-compatible online goal-reward paradigm. We propose—and experimentally validate—a novel decision-heuristic in which risk averse choice emerges from an inferential bias due to contingency neglect in the context of pairwise comparisons. We conclude with experimental evidence suggesting that this heuristic offers a potential explanation for empirical puzzles in other risky-choice settings of economic interest such as deductible-based insurance and portfolio allocation.

Acknowledgements: We extend a special thanks to Tim Houlihan and George Loewenstein for their insights and for facilitating access to these data. We additionally thank Linda Babcock, Karna Basu, Lynn Conell-Price, Stefano DellaVigna, Russell Golman, Kareem Haggag, David Huffman, Alex Imas, Botond Koszegi, Ted O'Donoghue, Alex Reese-Jones, 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 attitudes should inform how policymakers regulate asset and insurance markets, how firms design employee contracts, how individuals allocate their investment portfolios, and how economists evaluate the welfare consequences of policies and programs. From the perspective of expected utility theory (EU), the dominant framework in Economics for understanding decisions under risk, risk aversion, among fully informed, utility-maximizing individuals reflects 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 uncertainty) across a range of economic outcomes, however, gives rise to ostensible puzzles relative to EU (or the standard model).[1] A first puzzle is that people often exhibit a degree of risk aversion inconsistent with, and often in excess of, EU predictions (see Barseghyan et al. 2018). In the lab, researchers have documented a degree of risk aversion in small-to-medium sized gambles (e.g., Holt and Laury 2002) that implies implausibly high risk aversion at larger scales (Rabin 2000). And in the field, researchers have catalogued seemingly excessive risk-aversion in settings such as deductible-based insurance and portfolio choice.[2] A second puzzle, notably found in research on insurance demand, is that observed variation in perceived risk, or even preferences for risk, do not fully explain observed heterogeneity in risky behavior.[3]

Over the last few decades, researchers have advanced departures from the standard model to explain risk aversion through channels beyond diminishing marginal utility such as biased beliefs, non-linear decision weights (e.g., Kahneman and Tversky 1979; Prelec 1998), and loss/disappointment aversion in the context of gain-loss preferences (e.g., Kahneman and Tversky 1979; Koszegi and Rabin 2006; Gul 1991; Loomes and Sugden 1986). As illustration, risk aversion could reflect the overestimation of underlying risk, disproportionate weighting of unlikely outcomes, or an aversion to unanticipated losses. Risk aversion could also emerge from non-standard processes receiving less attention in

---

[1] For simplicity, we largely elide the 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] For example, several papers have documented the inconsistency between consumer demand for insurance and predictions of EU benchmarks (e.g., Barseghyan et al. 2013; Abaluck and Gruber 2011; Sydnor 2010; Handel 2013; Bhargava et al. 2017). In the context of deductible-based insurance involving property/health (i.e., high-probability, low-consequence insurance), papers routinely document excess demand for coverage relative to benchmark predictions (e.g., Sydnor 2010; Barseghyan et al. 2013).

[3] Researchers have found that heterogeneity in risk type does not explain variation in insurance demand (Cohen and Einav, 2007). Others have asserted neither heterogeneity in risk nor in risk preferences can explain variation in insurance demand (e.g., Cutler and Zeckhauser, 2004; Barseghyan et al. 2013). In an analysis of choice among experimental participants, Jaspersen, Ragin, Sydnor (2022) find only a modest correlation between lab-based measures of risk attitudes and insurance demand.

economics such as those involving heuristics, salience, affect, cognitive processes, or hormones (see Kusev et al., 2017; Fox, Erner, and Walters, 2015). Practically, clarifying motives for risk-taking in the field has been complicated by limited data on beliefs (e.g., betting markets, insurance), complexity of the choice environment (e.g., insurance, asset choice), or limited generalizability (e.g., game shows).

We overcome 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 firm specializing in the design and administration of programs leveraging 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 rewards increased non-linearly (e.g., Goal 1: $100, Goal 2: $300, Goal 3: $600). Much like other economic decisions under risk and uncertainty, one can interpret goal choice as a decision between simple lotteries varying in subjective risk and reward.

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 on goal choice and productivity, but not beliefs, yielding a composite sample of 35,478 employees with potential earnings of $17.5 million. To organize potential explanations for goal choice, we outline a simple framework in which a risk-neutral, fully-informed, employee selects a high or low goal with an all-or-nothing reward by maximizing expected utility. We then successively introduce departures from this baseline—preference-based risk aversion, biased beliefs, non-linear decision weights, and gain-loss utility—to generate alternative benchmarks from which to assess 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.

Several features of this setting position it as an attractive litmus test for understanding how people engage financial risk. First, the diversity of the sample, the near-complete participation rate, and the wide-ranging financial stakes enhance the generalizability of the sample. For example, 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. 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 variability in rewards, which range from $69 to $4,500—an interval that encompasses many financial decisions of interest to economists. Second, in contrast to settings where a lack of understanding of the decision environment may confound attempts to assess risk, GQ asks employees to make decisions from a standardized and simple choice menu. Finally, and most importantly, our partnership with BIW led to the development of an enhanced enrollment process through which we elicited contemporaneous employee beliefs of goal attainment. Access to decision-maker beliefs, unusual for field data on financial risk-taking, allowed us to directly test belief-based explanations for the observed behavior.

Our analyses of employee decisions and beliefs yield three main findings. A first finding is to document substantial risk aversion and choice diversity in the decisions of employees (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, 49 percent of employees selected a lower goal than predicted by the expected-value (EV) maximizing benchmark (for most employees, Goal 3). Conservative goal choice, for those attaining the low-goal threshold, resulted in an average loss in counterfactual reward of 46 percent ($164 compared to $303) relative to ex ante optimal choice. Overall, 45 percent of employees chose optimally with respect to the EV criterion, a share that did not meaningfully vary across reward size, employee tenure, or approximate salary (implying that risk-averse choice was not due to financial illiquidity).

Our second finding is that utility-based preferences for risk cannot explain conservative choice. Specifically, we show that allowing for a plausible degree of risk aversion, modeled as any CARA utility in the interval, $r \in [0.0003, 0.005]$, does not increase the explanatory power of the benchmark (we show this result persists under the assumption of CRRA utility). 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 relative to the risk- neutral benchmark, it increases the share of seemingly aggressive choice by a roughly offsetting magnitude. Even a benchmark of heterogeneous risk preferences characterizing choice as optimal if it can be rationalized by *any* risk preference in the plausible interval fails to explain over 40 percent of employee decisions.

Our third finding is that departures from EU—biased beliefs, non-linear decision weights, and gain-loss utility—routinely contemplated as alternative explanations for risk aversion cannot explain observed choice. For example, while a systematic bias in beliefs favoring low 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 high goal attainment. Indeed, a benchmark model of subjective expected utility with plausible risk aversion (SEU) explains only one-half of employee choices, a rate of accuracy only modestly exceeding the rational expectation benchmark. Similarly, while conservative choice could result from the assumption of non-linear decision weights, incorporating the popular weighting function of Prelec (1998) does not improve explanatory power. Finally, given the precedent in the literature for explaining conservative choice through the assumption of loss or disappointment aversion, we constructed a portfolio of benchmark models with gain-loss utility 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 improve explanatory power—the most successful among them accurately predicted 59 percent of choice.

To generate additional evidence on mechanisms and to rule out potential confounds, we administered an experimental incentive-compatible rewards program, resembling GQ, in the context of an online effort task. The experimental paradigm permitted us to observe multiple goal choices per participant in a setting where we could confirm understanding of program rules, explicitly denominate rewards in dollars, and minimize motives pertaining to reputation, signaling and costs of effort. 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 the full set of six elicited goal choices for more than one-quarter of participants. We additionally tested—and found no evidence to support—alternative explanations such as choice heuristics in which people sort themselves into options from an ordered menu based on contextually-relevant cues such as ability or taste for competition (Kamenica 2008; Niederle and Vesterlund 2007).

We conclude by proposing a novel heuristic to help explain the documented behavior in the lab and the field. Informed by open-ended descriptions of decision processes from pilot studies and the broader literature on decision-making, our proposed “pairwise heuristic” stipulates that an individual engages risky choice through a succession of approximate pairwise comparisons between proximal options. Critically, the pairwise comparisons lead to inferential errors associated with the propensity to neglect contingent probabilities. In the case of GQ, contingency neglect leads employees to systematically underestimate the relative likelihood of attaining riskier goals, prompting greater conservatism than predicted by unbiased benchmarks. As a concrete example, the heuristic implies an employee deciding between Goals 2 and 3 (having ruled out Goal 1) would roughly assess whether the expected potential gain from selecting the high goal, assuming attainment of the low goal (i.e., the difference in rewards weighted by the perceived conditional likelihood of high-goal attainment, $\Delta x_{3,2} * \hat{s}_{3|2}$), offsets the expected potential loss from not selecting the low goal ($\hat{s}_{\neg 3|2} * x_2$). The employee underestimates the conditional likelihood of goal attainment, however, due to insufficient adjustment for the contingent nature of the comparison ($\hat{s}_{h|l} = ks_{h|l}, k \in [s_l , 1)$). The heuristic predicts greater risk aversion—and choice diversity—than prior benchmarks. While the heuristic has not been previously discussed, it builds on conjectures from the literature such as biased relative evaluation (e.g., Koszegi and Szeidl 2013; Bushong, Rabin, and Schwartzstein 2021) or inference (see Benjamin 2019), salience and selective attention (Bordalo, Gennaioli, and Shleifer 2012), the failure to engage probabilistic (Sunstein 2002) and/or contingent (Martínez-Marquina, Niederle, and Vespa 2019; Sunstein and Zeckhauser 2010) information, and noisy decision-making (e.g., Camerer 1989; Hey and Orme 1994; Kahneman et al. 2021).

We sought evidence for the pairwise heuristic from a new experiment. The experiment—which asked hundreds of participants to make a hypothetical goal choices from one of several experimentally varying menus and queried decision-relevant beliefs—yielded several insights as to the plausibility of the heuristic. First, beyond corroborating the pattern of conservative and diverse choice from the field (relative to a risk-neutral SEU benchmark), the experiment affirmed participant use of proximal pairwise comparisons and the systematic, and often substantial, neglect of contingent likelihoods. Second, controlling for non-contingent beliefs of goal attainment, the magnitude of the within-subject inferential bias strongly predicted optimal choice. Lastly, when randomized to select a goal from a debiased menu—i.e., a menu displaying empirically accurate contingent likelihoods of goal attainment—participants were 48 percent more likely to select optimally than a menu displaying the equivalent non-contingent likelihoods. Moreover, participant response to a menu displaying no attainment probabilities (we provided participants context through a simulated performance history) was indistinguishable from response to a menu displaying contingent likelihoods adjusted for the hypothesized bias ($k = s_l$). As the final test of the heuristic, we assessed its capacity to predict observed goal choice in the lab and the field and found that it substantially outperformed prior benchmarks. With a conservative allowance for noise, the pairwise heuristic explained 54 to 60 percent more choice than the SEU benchmark in the lab and, despite an inability to observe employee-specific bias in beliefs, 26 to 46 percent more choice in the field. The evidence implies a moderate, to potentially very large, share of employees engaged in heuristic choice.

We speculate that because of the centrality of contingent evaluation for decisions under risk and uncertainty in settings such as insurance or asset allocation, the proposed heuristic may help resolve empirical anomalies in economic risk-taking more broadly. We conclude with a final experiment investigating the applicability of the pairwise heuristic for understanding deductible-based insurance, a choice domain not easily explained through standard benchmark models, even after modification for common behavioral departures (e.g., Jaspersen, Ragin, Sydnor, 2022). In this setting, our heuristic predicts, relative to other benchmarks, greater demand for coverage due to neglect of the non-focal contingency (i.e., the possibility of no covered losses), increased heterogeneity in choice due to underestimation of the relative likelihood of severe losses, and a diminished correspondence between perceived risk, risk preferences, and demand. We implemented the experiment by asking participants to select a home insurance plan from menus adapted from Sydnor (2010), an analysis documenting excess demand for coverage among real-life homeowners. Consistent with predictions, the experiment revealed substantially greater demand for the EV-maximizing plan, and less choice diversity, among participants selecting from a debiased menu encouraging consideration of the non-focal contingency—via the display of empirically-informed likelihoods of no damage—compared to a baseline menu with no display or menus encouraging relative evaluations within the focal contingency (via the display of empirically-informed relative likelihoods of less versus more severe damage).

Our research relates to multiple literatures in economics. First, we contribute to work aspiring to understand the extent of and motives for financial risk-taking in the field (see Barseghyan et al. 2018). Beyond documenting substantial risk aversion in an environment distinguished by decision simplicity, diversity of economic stakes, and data transparency, we leverage our access to decision-maker beliefs to explicitly reject—for roughly one-half the sample—predictions of the standard EU benchmark as well as popular alternative motives involving biased beliefs, non-linear decision weights, and gain-loss utility. Next, the explanation we do propose for employee choice concurs with other research emphasizing the potential importance of decision heuristics for understanding consumer behavior in domains such as insurance (e.g., Ericson and Starc 2012; Bhargava et al. 2017; Jaspersen, Ragin, Sydnor, 2022) or asset allocation (e.g., Benartzi and Thaler 2007). However, because the pairwise heuristic implies potentially substantial inferential error, our findings highlight the risks of sub-optimal design and biased welfare analyses in policy settings where policymakers and researchers mischaracterize underlying decision processes. As a concrete example, the lack of descriptive invariance in insurance demand from our experiment alludes both to the opportunity to improve welfare by strategically reframing choice and to the challenges of naively inferring risk attitudes from choice. Finally, the present research contributes to recent work recognizing the possibly underappreciated role of probability/contingency neglect for understanding household financial decisions (Martínez-Marquina, Niederle, and Vespa, 2019; Sunstein and Zeckhauser, 2010).

2 BACKGROUND

2.1 Institutional Background

GoalQuest® (GQ) is an employee-rewards program conceived and administered by BI WORLDWIDE (BIW), a private global consulting firm. The firm, founded in 1950, specializes in the design and delivery of a suite of proprietary programs that leverage principles from behavioral science (e.g., non-monetary rewards, goal-setting, personalization, symbolic recognition, lotteries, contests, communication, and feedback) to improve employee, channel partner, and consumer engagement. As of 2021, BIW had self-reportedly engaged 6 million individuals across 144 countries through its various products. As of the same date, according to third-party estimates, the firm had approximately 1,500 employees and annual revenues between $500 million to $1 billion. Described by BIW as the world's only patented incentive-based sales program, GQ was designed to motivate employee productivity through self-selected performance goals tied to all-or-nothing non-monetary rewards.[4] As of 2018, BIW had administered over 1,000 GQ programs to over 1 million participants at firms primarily in the United States, Canada, and Europe since its 2001 inception. While marketed as a sales incentive program, our data indicate that the program has serviced a significant share of employees engaged in customer service and retention (e.g., call centers) across a diversity of sectors (e.g., communication, health care, manufacturing, financials, consumer discretionary).

2.2 GQ Program Overview

Across the wide-range of client firms, GQ boasts a uniform program structure, particularly since 2012, the year of our earliest data. For the few hundred to a few thousand employee participants of a typical 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.[5] The webflow itself consists of three parts: a program overview, an enumeration of program rules, and goal selection (see Appendix for select screenshots). Employees select a goal from a menu of three personalized options (Goal 1, Goal 2, Goal 3) each associated with an all-or-nothing reward denominated in points.[6] Speaking to generalizability, BIW promotes the program as having a 98 or 99 percent participation rate among eligible employees.[7]

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

[4] The World Intellectual Property Organization Publication Number associated with GQ is WO 01/13306 A2 (February 2001).
[5] While the design of the enrollment portal was standardized across programs at any point in time, its design evolved over time.
[6] In some programs, the goal selection period may have briefly overlapped the performance window.
[7] While we cannot directly verify participation statistics, high participation rates are plausible due to marketing and communication during the pre-period, the administrative ease of enrollment, and often-valuable rewards.

achievement levels?" (scale indexed in increments of 10 percentage points). Employees were additionally asked about their gender, age, and tenure with the firm. Presumably due to its integration within the enrollment process, while technically optional, survey participation across our sample was 60 percent.

Following goal selection, employees proceeded to the several-week performance period during which they attempted to achieve their selected goal. In most programs, participants were able to log onto the website to check their progress or to remind themselves of their selected goal.[8] At the close of the performance period, employees who attained their goal exchanged reward points for a reward in the GQ marketplace. The non-monetary rewards included major electronics (e.g., 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 during program marketing; for many programs, employees were familiar with the marketplace through other BIW programs using the same point currency.

2.3 Goal and Reward Structure

Major elements of the GQ goal and reward structure were designed to increase employee productivity, premised on research in behavioral science.[9] First, based on the presumed importance of personal choice and personalization, GQ required employees to self-select a goal from a menu personalized from the employee's prior productivity. 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).[10] 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 usually segregated into a few 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 adopt different rules for employees who differed in their baseline performance. Employees were not given any guidance during the goal-selection process via recommendations, defaults, or persuasion.

---

[8] 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.

[9] 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).

[10] The calculation of baseline performance was jointly determined by BIW and each firm on a program (and often group) specific basis based on considerations of data availability, employee tenure, and seasonal variation in productivity. For many programs, the baseline was calculated from employee performance over a recent period of similar duration to the program.

Second, based on the presumed motivational potency of high goals, GQ implicitly encouraged higher goal choice by engineering them to be more financially attractive, for most employees, than other goals, in expectation. Specifically, in contrast to the additively linear increase in goals, rewards typically increased in non-linear increments. For example, many reward menus followed the *k*, 3*k*, 6*k* structure, where *k* was set to be approximately 1 percent of an average employee's salary over the course of the program. Moreover, goals 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 monetary rewards of similar value.

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 the likelihood of goal attainment. An employee must select exactly one of the two lotteries. We describe an employee's expected valuation of a goal-lottery as: $\mathrm{V}(G_n) = \pi(\hat{s}_n)\, v(x_n, \theta)$. Here, π denotes a decision-weighting function applied to some subjective probability of goal attainment and $v(.)$ is an always increasing function, potentially dependent on a reference point, θ, denoting an employee's preference for rewards.

An employee will choose the low goal if $\pi(\hat{s}_l)v(x_l,.) \geq \pi(\hat{s}_h)v(x_h,.)$. The framework helps to identify potential reasons for conservative goal choice including utility-based risk aversion, biased beliefs ($\hat{s}_n \neq s_n$), non-linear decision weights ($\pi(\widehat{s_n}) \neq \widehat{s_n}$), and gain-loss utility.

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, where *r* captures an employee's attitude towards risk (i.e., $r > 0$ implies risk aversion, $r = 0$ denotes risk neutrality; we ignore the possibility of $r < 0$):

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

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 (CRRA) across varying 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.

Risk Neutrality ($r = 0$). 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.

Risk Aversion ($r > 0$). 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. 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 [$\hat{s}_n \neq E(s_n)$]

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 bias in beliefs. As a result, $\gamma_n > 1$ implies overconfidence while $\gamma_n < 1$ implies underconfidence. A risk averse employee with biased beliefs will select the low goal if:

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

The decision rule implies that conservative goal choice increases in γ_l/γ_h. For example, if employees were systematically under-confident about future productivity, and this led employees to deflate the likelihood of achieving higher, relative to lower, goals, then one might expect utility-maximizing employees to act more conservatively than predicted by the benchmark ($\gamma_l/\gamma_h > 1$). Alternatively, employee overconfidence could generate conservative choice if such overconfidence led to employees to systematically inflate the likelihood of achieving lower, relative to higher, goals (also, $\gamma_l/\gamma_h > 1$).

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. 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):

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

In theory, if employees systematically underweight moderate-probability outcomes (e.g., Goal 3) relative to higher-probability outcomes (e.g., Goals 1 and 2), a non-linear weighting function might help explain goal choice. 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 $[v(.) = v(x_n, \theta)]$

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 a range of economic contexts. Given that the structure of GQ stipulates that every employee receives either nothing or 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).

One practical challenge for assessing models of gain-loss preferences, however, is the absence of clear theoretical guidance as to how to represent gain-loss utility. In service of assessing the breadth of plausible representations, we appeal to the theoretical literature, and practical considerations of the decision context to consider candidate reference points, functional specifications, and loss aversion magnitudes. Regarding the former, 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. 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. We further consider two prominent functional representations of gain-loss utility—gain-loss utility in isolation and a composite framework 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 consumption utility and gain-loss utility are additively separable and specify η as a scaling factor applied to consumption utility so that η = 0 reduces to a model with gain-loss utility only.

All considered, we represent gain-loss utility, given some reference point, θ, as follows:

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

The term $m(x_n)$ can be thought as a classical utility function that is strictly increasing, utility over gains, u^{+}, is concave, while utility over losses, u^{-}, is convex. We evaluate this function for a range of potential reference points, loss aversion parameters, and scaling factor values.

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 describes demographic detail, goal choice, employee productivity, and employee beliefs of goal attainment. The program-level data describes the identity of 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 central evidence on employee behavior and beliefs draws from what we denote as a *primary sample*. This sample comprises 20,133 decisions and corresponding beliefs constructed by applying screening restrictions to an original dataset from BIW. This original data, which spans 38,661 employees across 34 programs and 18 firms, reflects 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 employees with full participation, and whose data had been electronically archived by BIW.[11] From this original data, we first generated an *expansive sample* (n = 35,478) by excluding roughly 8 percent of records for which a key data field was missing (excluding employee salary for which we only have partial coverage), the data was inconsistent, or we inferred the employee did not complete the program.[12] To create the primary sample, we then restricted the expansive sample to employees who provided internally consistent beliefs

[11] Data for a small number of programs was not archived by BIW. The size cutoff was practically necessitated by the burdens of organizing and transferring program data (BIW) and the resources required to audit every employee in each program.
[12] 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.

in enhanced enrollment.[13] 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 may slightly underestimate the degree of conservatism and sub-optimal choice in the broader population of employees.[14] We reproduce key analyses for the expansive sample in Section 5.

Table 1 summarizes overall sample statistics as well as group-level (duration, financial stakes) and employee-level (age, gender, tenure, inferred income) characteristics for the 20,133 employees across 18 firms, 34 programs, and 232 distinct groups in the primary sample.[15] On average, we observe data for 592 employee participants per program (IQR: 208 to 703) and 87 employees (IQR: 12 to 103) per group. The groups varied with approximate uniformity across either 30, 60, and 90-day program durations (two programs ran for 45 and 120 days). The distribution of potential reward values was asymmetric, such that 10 percent of employees engaged decisions with rewards averaging $2,150, despite a group-level average of $607 and an employee-level average was $466. Overall, employees in the primary sample could have earned up to $9.4 million in possible rewards, while those in the expansive sample could have earned $17.5 million in potential rewards. The table also conveys the diversity of the sample across gender, age, and tenure. We suspect that the average program-level salary of $70,400 overstates the average salary across the sample because the 8 of 18 firms for which we observe salary feature higher rewards than firms for which we do not observe salary (and average reward values correlate to pro-rated salary).

4.2 Goal Choice, Employee Productivity, and Goal Attainment

Our analysis relies on measures of goal choice, productivity, and beliefs. We describe goal choice, g, both through indicators of choice for each of the three goals and indicators characterizing the optimality, aggressiveness, or conservativeness of each goal. To characterize choice, we compare the expected utility of the selected goal with non-selected goals using the benchmark models outlined in the theoretical framework. We describe an employee's productivity with two normalized measures: (i) productivity relative to baseline, and (ii) productivity relative to the Goal 3 threshold.[16] Normalization

[13] 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.

[14] 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).

[15] 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.

[16] We did not have baseline data for 16 percent of employees. In most cases, this reflects the lack of past performance data for new employees or programs where performance goals were defined without reference to a baseline.

permits comparisons across programs where productivity varies in scale (e.g., productivity may be measured in hundreds of calls at a call center; it may be measured in single-digit sales at an electronics firm). Finally, for each employee, we calculate indicators of baseline and goal attainment.

Table 2 summarizes employee choice, productivity, and attainment. The table indicates that 44 percent of employees selected the highest goal with a roughly even split across the other two goals. The table indicates a correlation between goal choice and productivity, consistent with more productive employees sorting themselves into higher goals (or possibly that higher goal choice led to elevated performance). The table also shows that while only 29 percent of employees attained the highest goal, 66 percent of those attaining the lowest goal also attained the highest goal (i.e., 0.29/0.44).

4.3 Employee Beliefs

We calculate two measures of employee beliefs—subjective beliefs elicited through enhanced enrollment and estimates of ex ante rational expectations. Our measure of employee *i*'s subjective belief of attaining goal k, $\hat{s}_{k,i}$, simply reflects the employee's response to the interval-elicitation of belief from enhanced enrollment. To estimate an employee's rational expectation of goal attainment, $\hat{s}_{k,i}^{r}$, we appeal to a strategy routinely used in research on insurance. First, we segregated employees by program group and goal choice. Next, for each employee, we predicted the ex ante likelihood of goal attainment for each goal by adjusting the group x goal average by observable covariates. The exercise effectively assumes that one can proxy for rational expectations with the average attainment of similar others, a strategy commonly used in the economic literature on insurance. More specifically, we initially estimated the following leave-out regressions for each employee *i* and goal $k \in [1,2,3]$: $\bar{s}_{k,l,-i} = \alpha + \boldsymbol{Z}\gamma + \pi_l + \varepsilon$. Each regression predicts average group-level attainment for each goal, $\bar{s}_{k,l,-i}$, leaving out employee i, as a function of employee characteristics included in vector **Z** (age, tenure, gender) and group fixed effects, π_l. (We estimated regressions at the program level to increase the precision of covariate estimates). We then calculated an employee's rational expectation of attaining goal k, as $\hat{s}_{k,i}^{r} = \hat{\alpha} + \boldsymbol{Z}\hat{\gamma} + \hat{\pi}$.

Table 3 summarizes data on employee beliefs and highlights two notable patterns. First, the table documents a correlation between expectations of attainment and goal choice. Second, the comparison between subjective and rational beliefs suggests substantial overconfidence among employees regarding future productivity. Employees exhibited significant overconfidence, on average, with respect to every goal. And importantly, such overconfidence was greater for higher, relative to lower, goals.

5 CHARACTERIZATION OF GOAL CHOICE BY BENCHMARK MODEL

We now characterize employee choice relative to predictions of the benchmark models outlined in the theoretical framework. Tables 4 and 5 report the share of optimal (goals matching benchmark prediction), conservative (goals lower than benchmark prediction), and, for completeness, aggressive (goals higher than benchmark prediction), choice for the various benchmarks. To understand the economic consequences of sub-optimal choice, we also report the average expected and unrealized gain for different choice types. Finally, to better understand the moderating role of financial stakes and experience, we report optimal choice shares across reward size quartile and categories of tenure.

5.1 Expected Utility with Risk Neutrality

We first assess choice using a baseline benchmark that assumes risk-neutral employees with rational expectations select goals by maximizing expected utility before modifying the benchmark with subjective beliefs. As presented in Table 4, under the rational expectation benchmark, 45 percent of goal choices are characterized as optimal while most remaining choice is characterized as conservative. For employees who chose conservatively, and attained at least some goal, the actual realized gain of $164 relative to the average potential gain of $303 under ex ante optimal choice implies a counterfactual loss of 85 percent. The share of optimal choice did not vary across reward size or employee tenure. Figure 1, which depicts the cumulative distribution of counterfactual loss overall and by goal choice for those achieving Goal 1, indicates the concentration of loss among those selecting one of the two lower goals.

Replacing rational expectations of goal attainment with subjective beliefs did not meaningfully reduce the share of conservative choice and led to a small improvement in the share of choices characterized as optimal. The increase in optimality under subjective beliefs is largely due to the reclassification of most previously aggressive choices, under rational expectations, to optimal (the number of previously conservative choices reclassified as optimal was nearly offset by the number of previously optimal choices reclassified as conservative). Adopting subjective beliefs did not shift the moderation of choice optimality by reward size or tenure. Table 3 provides additional insight as to how replacing rational expectations with self-reported employee beliefs failed to 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. The table conveys substantial employee overconfidence with respect to attaining each of the three goals and substantial relative overconfidence about attaining higher, versus lower, goals. While relative overconfidence increased the gap in expected value between lower goals and Goal 3, it only increased the share of employees for whom Goal 3 was ex ante optimal from 84 to 87 percent, since Goal 3 was attractive in financial expectations for most employees under rational expectations.

5.2 Expected Utility with Risk Aversion

We now consider the possibility that conservative choice may reflect risk aversion attributable to the diminishing marginal utility of wealth. We model risk aversion by assuming a CARA utility function with any plausible risk aversion, $r \in [0.0003, 0.005]$. To appreciate the breadth of risk attitudes captured by this interval, we can translate what such risk preferences imply for gambles involving potential losses of a size comparable to the rewards available 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 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 would accept any 50/50 gamble involving a potential loss of \$350 (the median GQ reward value) so long as the potential gain exceeds \$391. The other endpoint, $r = 0.005$, implies an employee would reject *any* 50/50 gamble involving a potential loss of \$175 (or \$350), even if the potential gain was infinite. In this sense, $r = 0.005$ offers a highly conservative upper bound of plausible risk aversion for financial gambles in the range of interest.

Table 4 characterizes choice for expected utility maximizing employees under either rational or subjective expectations assuming plausible risk aversion. 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. The assumption of more severe, but arguably plausible, risk aversion ($r = 0.005$) moderately shifts the characterization by reducing the share of conservative choice offset by an increase in the share of seemingly aggressive choice. Critically, assuming severe risk aversion does not substantially shift the overall descriptive accuracy of the expected utility benchmark, the share of counterfactual loss associated with conservative choice, nor moderation in choice optimality by reward size or tenure, under either information regime. Figure 2, which plots the share of optimal and conservative choice under the EU benchmark across different assumptions of beliefs, graphically depicts the relative insensitivity of choice characterization to the size of the potential reward.

Figure 3 helps conveys the intuition for how the assumption of preference-based risk aversion affects the characterization of optimal choice. The figure depicts the share of optimal goal choice for the expected utility benchmark for r ranging from $r = 0$ to $r = 0.10$ by goal choice assuming either rational expectations (Panel A) or subjective beliefs (Panel B). Across panels, the figure shows that increasing the degree of assumed risk aversion, within the plausible range, modestly increases the optimality of those selecting Goals 1 and 2 but reduces the optimality of employees selecting Goal 3 by an offsetting degree. In the Appendix we recharacterize goal choice for benchmarks with CRRA utility, assuming either rational or subjective beliefs, across a range of potential wealth levels and degrees of relative risk aversion. The results, summarized in Appendix Table A2, indicate that the assumption of CRRA utility, for plausible relative risk aversion, yield choice characterizations nearly identical to CARA benchmarks.

<u>Heterogeneous Risk Preferences</u>. 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 $[0, 0.005]$ for the expected utility benchmark under either rational or subjective beliefs. As reported in the table, shifting from a benchmark model assuming severe, but uniform, risk aversion, to one assuming heterogeneous risk preferences increases the share of optimal goal choice from 0.44 to 0.56 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 revisit the possibility that decisions may reflect diversity in risk preferences in subsequent experimental analyses.

Characterizing Choice in Expansive Sample. We replicate the preceding analysis for the expansive sample (i.e., the sample inclusive of the primary sample and the additional 15,345 employees for whom we do not have data on beliefs). Overall, employees excluded from the primary sample appeared to choose even more conservatively than those in the primary sample despite nearly identical productivity. Specifically, from the perspective of the expected-utility benchmark with rational expectations, the characterization of choice with the expansive sample resembled that of the primary sample but for a larger share of conservative choice and a smaller share of optimal choice (0.41 optimal; 0.56 conservative; 0.04 aggressive).[17] As with the primary sample, the introduction of moderate to severe CARA risk aversion to the benchmark did not meaningfully influence the optimal choice share; assuming severe risk aversion (r = 0.005) shifted the classification of some conservative choice to aggressive choice (0.40 optimal; 0.48 conservative; 0.11 aggressive). The analysis of the expansive sample suggests that to the extent the primary sample isn't fully representative of the employee population, it underestimates the degree of sub-optimal and risk-averse choice with respect to standard benchmarks.

5.3 Non-Linear Decision Weights and Noisy Beliefs

We proceed to consider whether non-linear decision weights or an allowance for noisy beliefs may help to explain employee choices. Specifically, to assess the former, we replace the linear decision weights assumed by the subjective expected utility benchmark (r = 0.0003) with the popular inverse s-shaped weighting function suggested by Prelec (1998; α = β = 0.65). Table 5 indicates that the modified benchmark does not meaningfully shift choice characterization relative to baseline. The absence of change in characterization is perhaps unsurprising given that the effective shift in weights associated with typical non-linear weighting functions are not pronounced for moderate to high beliefs.

We also consider whether modifying the benchmark to accommodate potential noisiness in beliefs, either due to actual uncertainty in such beliefs or uncertainty associated with our elicitation

[17] To characterize choice under rational expectations in the expansive sample, we adhere to the previously adopted strategy but for excluding unobserved demographic variables in the regression estimates of beliefs.

procedure. We implement an allowance for noise by evaluating whether the subjective EU model with moderate risk aversion can rationalize choice for any set of subjective beliefs within a +/- 10 percent range of the self-reported figure. As indicated in the table, the allowance of a 20 percent error in subjective beliefs only moderately increases the share of choice deemed to be optimal, from 0.50 to 0.54.

5.4 Gain-Loss Utility

Finally, we consider the possibility that conservative choice may reflect prospective loss aversion in the context of gain-loss utility. In accordance with our interest in testing all credible representations of gain-loss utility, we evaluate a large combination of benchmark models reflecting varying candidate reference points, θ, functional scaling factors, η, and loss aversion parameters, λ, as specified in the theoretical framework. Specifically, we 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 additionally considered prospect-dependent reference points including the chosen goal, the expected value of the chosen goal, and in recognition of models of counterfactual regret, the nearest-goal either below or above the chosen goal. We assessed these reference points in the context of several potential composite utility specifications generated by assuming a KT power function ($\alpha = 0.88$) for both consumption and gain-loss utility components across values of η ranging from 0 to 5 (a necessarily wide range given the lack of empirical consensus in the literature as to the appropriate weighting). Lastly, in deference to the breadth of loss aversion parameters contemplated by the literature, we consider loss aversion parameters of $\lambda = 1.5$, 2.25, and 3.0. To streamline the analysis, we assume linear decision weights and subjective employee beliefs.

Table A1 of the appendix reports the descriptive accuracy of the gain-loss benchmarks (a red-to-green gradient helps interpret relative model efficacy). Among the prospect-independent reference points, nearly all explain approximately one-half of all goal choices, a rate of descriptive accuracy not different from the previously considered benchmarks. Among prospect-dependent reference points, the benchmarks exhibit more variation in their descriptive accuracy, largely driven by variation in reference points and the scaling parameter. Across all tested benchmarks, the most successful, with a reference point set at the chosen goal reward ($\eta = 1$, $\lambda = 2.25$) explained 59 percent of employee choices.

Table 5 reproduces the full characterization of choice for the most promising benchmark with gain-loss utility alongside other non-standard benchmarks. Beyond delivering a moderate increase in explanatory power relative to previous benchmarks, the table shows that the gain-loss benchmark does not exhibit moderation in descriptive accuracy by reward size or tenure. Collectively, our attempts to characterize choice with expected-utility benchmark models, even allowing for common behavioral departures, failed to explain over 40 percent of employee choices with little moderation by stakes and experience and persistently implied a substantial degree of conservative choice. We explore additional explanations, and attempt to rule out potential confounds, for the observed conservatism through experiments using an online goal choice paradigm.

6 EXPLORING MECHANISMS VIA EXPERIMENTS

We further investigate the motives for conservative goal choice through two online experiments. The first experiment was intended to corroborate findings from the field with increased statistical power, rule out potential confounds, and assess alternative explanations from the literature. The second experiment 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 a brief effort task in the context of an incentive-compatible goal-reward paradigm. The paradigm resembled GQ but with lower but dollar-denominated stakes, a shorter evaluation period, comprehension checks, and multiple decisions per subject. We supplemented the paradigm with decision-relevant questions including an elicitation of beliefs, assessments of risk and loss aversion, and self-assessments of relative ability and taste for competition.[18]

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. To solve a grid, participants had to find the unique pair of numbers whose sum equaled 10 within a 3 x 3 matrix of single-digit numbers (the task resembled those previously used in the literature). After an opportunity to practice solving grids, we formally introduced participants to the goal-reward paradigm, which we named GoalQuest, via an online webflow resembling that used in field. The webflow explained to participants that they 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 a series of questions to test comprehension of the paradigm, participants proceeded to goal selection.

To increase our 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, presented one-by-one, strategically varied the spacing of the goals and rewards as well as the

[18] 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.

number of options in order to facilitate tests of mechanisms (as informed by various pilot tests). 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 additional menus varied either overall difficulty or the financial attractiveness of Goal 3 relative to the other goals, and two additional menus expanded the baseline menu by adding a relatively unattractive high- or low-goal option. After participants selected their goals, we elicited performance expectations by asking them 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).[19] We additionally asked participants to forecast how many grids they expected to complete. Finally, participants completed the four-minute effort task and were awarded a bonus based on goal attainment from the randomly selected goal menu.

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 resembling employees in the field. Average baseline choice across the three goals (Goal 1, 2, 3) from the lab (0.34, 0.28, 0.38) approximated choice in the field (0.29, 0.27, 0.44) as did beliefs of goal attainment from the lab (0.80, 0.66, 0.51) and field (0.78, 0.69, 0.63). Participants in the lab also exhibited overconfidence 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). Finally, using the risk-neutral subjective EU benchmark, baseline choice characterization (optimal, aggressive) in the lab (0.50, 0.45) was highly similar to the field (0.50, 0.48).

We interpret the correspondence in choice, beliefs, and particularly choice characterization across the lab and field as evidence discounting potential confounds involving program confusion, managerial signaling, or reputational concerns (while the latter two motives would presumably nudge employees towards more aggressive, rather than conservative, 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 minimal scope for signaling or reputational concerns.[20] Appendix Table A3 summarizes choice, beliefs, and attainment for 3-option menus from the experiment.

---

[19] 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.

[20] A possibility is that conservative goal choice in the field largely reflects employee intent to pre-emptively commit themselves to easier goals to avoid the perceived effort costs associated with ambitious goals. While possible, we see this explanation as

Results - Characterization of Goal Choice. Table 6 summarizes the optimality of choice relative to a range of benchmark models.[21] Beyond reporting the participant share whose full set of choices conform to the predictions of a particular benchmark, the table also characterizes choice after allowing for error in the form of a participants whose choices mostly adhered to the benchmark. The table indicates that previously considered benchmarks can explain at most 24 percent of participant choices, a rate that rises to 42 percent when allowing for error in the form of at least 5 of 6 optimal choices.

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 rational strategy for someone unsure of what goal to select but who believed the menu was 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 uninformed consumer decisions from product menus (Kamenica 2008). 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.[22]

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 a series of exploratory pilot studies in which we asked participants to describe the details of their choice deliberations and by 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 contingency-specific pairwise comparisons between proximal options. Importantly, the pairwise comparisons are associated with inferential errors involving decision-maker neglect of the contingency. In the context of the GQ menu, the heuristic would lead an employee to underestimate the likelihood of riskier events, increasing the likelihood of conservative choice.

---

unlikely given that (a) our data describing goal attainment beliefs were collected after employees selected their goal, (b) many conservative choices were made by employees who perceived high-goal attainment as very likely, (c) we find a similar pattern of conservative choice in the lab where effort-motives should be diminished.

[21] 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.

[22] 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.

We outline the heuristic more formally by returning to our earlier theoretical framework and notation where we represent goal choice as a decision from a menu of two binary gambles ordered from low to high risk (G_l, G_h). For simplicity, we focus on the decision of a risk-neutral employee and assume linear decision weights. The heuristic specifies the employee will evaluate the two options by assessing whether the expected potential gain from shifting from the low to high goal exceeds the expected potential cost from such a shift, allowing for some computational error. More concretely, the heuristic implies the employee will select the high goal if the expected potential increase in reward, assuming low-goal attainment, exceeds the expected potential loss of the low-goal reward, weighted by the likelihood of low-goal attainment:

$$\hat{s}_l\left[\left(\hat{s}_{h|l} * \Delta x_{h,l}\right) - \left(\hat{s}_{\neg h|l} * x_l\right) + \varphi\right] > 0$$

Here, the parameter $\hat{s}_{h|l}$ denotes the employee's 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 given low-goal attainment, $\Delta x_{h,l}$ is the gain in rewards from attaining the high, relative to the low, goal, and x_l denotes the low-goal reward. Since goal choice is irrelevant in this context if an employee fails to attain the low goal, she can simplify the evaluation to the following: $\left(\hat{s}_{h|l} * \Delta x_{h,l}\right) - \left(\hat{s}_{\neg h|l} * x_l\right) + \varphi > 0$.

If employees had unbiased and otherwise well-calibrated beliefs about goal attainment, then the pairwise comparison simply restates the utility-maximizing proposition, excepting the noise allowance. The heuristic, however, dictates that an employee systematically neglects contingent probabilities in the relative evaluation of options, leading the employee to systematically underestimate the financial value of switching from the low to high goal and subsequently to more conservative choice. Assuming an employee partially to fully neglects to adjust the posterior likelihood of high-goal attainment by the marginal probability of low-goal attainment, we can represent the bias as $\hat{s}_{h|l} = k s_{h|l} = k \frac{s_h}{s_l}$, where k $\in [s_l\ , 1)$. Notably, the heuristic also predicts an employee's evaluation would neglect adjustment by $\hat{s}_l$ but such neglect is inconsequential in the case of GQ.

To illustrate how the heuristic might affect goal choice, consider the stylized example in which an individual must select between a simple low-risk lottery with a reward of $300 and a success likelihood of 0.75, G_l = ($300; 0.75) and simple a high-risk lottery, G_h = ($600; 0.50). To more closely parallel GQ, assume that lotteries are realized through the same random process, a single draw from a uniform distribution [0,1] where G_l pays for [0.25, 1.00] and G_h pays for [0.50, 1.00], such that $s_{h|l}$ = 0.67. A risk-neutral individual who maximizes expected utility free from bias or imprecision would select the high goal lottery given that the marginal expected gain from the additional risk of the high goal, $200 ($300 x 0.67) exceeds the marginal expected loss of $100 ($300*0.33). In contrast, an individual governed by the pairwise heuristic, with full contingency neglect of the form $\hat{s}_{h|l} = s_h$, might select the low goal lottery, depending on the value of the noise parameter, φ, since the marginal expected gain from the high goal, $150 ($300 x 0.50), is now equivalent to the marginal potential loss ($300 x 0.50). In this way, the inferential bias, with computational imprecision, could help to explain lower goal choice.

While we have so far restricted discussion to a menu of two options, decision-makers could apply the heuristic to moderately larger menus through one of several strategies. For tractability, we assume that for with menus, such as GQ, with three risk-ordered options, employees apply the heuristic by successively comparing proximal pairs of options beginning at the low-risk option and stopping any time a riskier option is rejected. As such, we assume an employee would initially compare Goals 1 and 2, and either accept Goal 1 or proceed to Goal 2 (alternatively, for example, one could assume that decision-makers evaluate all possible pairwise comparisons).[23]

<u>Motivating Evidence from the Literature</u>. Three key assumptions about decision-making underlie the proposed heuristic: (i) the asserted use of pairwise comparisons, (ii) systematic bias in contingent inference, and (iii) the allowance for computational error. These assumptions draw from an extensive interdisciplinary literature. For example, the propensity of individuals to engage in relative, or comparative, evaluation is an established tenet in the study of decision-making with considerable experimental and neuroscientific support. Within economics, the comparisons are central to models of reference-dependent utility (e.g., Kahneman and Tversky 1979; 1992; Koszegi and Rabin 2007) and a growing literature contemplating the role of relative thinking in how people evaluate different dimensions of consumption (e.g., Bordalo, Gennaioli, and Shleifer 2012, 2013; Koszegi and Szeidl 2013; Bushong, Rabin, and Schwartzstein 2021). The proposed bias in contingent inference is in the spirit of economic/ statistical models of inferential bias involving systematic Bayesian departures (see Benjamin 2019), but perhaps more directly relates to discussions of probability/contingency neglect (Sunstein 2002; Martínez-Marquina, Niederle, and Vespa 2019; Sunstein and Zeckhauser 2010). Finally, the allowance for approximation or noise in decision-processes is a recognized component of decision-making frameworks both outside of and within economics (e.g., Camerer 1989; Hey and Orme 1994; Kahneman et al. 2021).

6.3. Experimental Evidence for Pairwise Heuristic (Experiment B)

We assess the plausibility of the proposed heuristic through two strategies. First, we present evidence from an experiment designed to test whether individuals adopt the two key decision-making precepts underlying the pairwise heuristic—i.e., reliance on proximal pairwise comparisons and underestimation of pairwise contingent probabilities—whether the magnitude of inferential bias predicts

[23] Practically, in even larger menus, where decision-makers are unlikely to consider all options, we speculate that they may apply the heuristic to a focal subset of options determined.

goal choice beyond beliefs alone, and whether a de-biased menu that discourages contingent evaluation leads to more optimal choice, as specified by standard benchmarks. Second, we assess whether the pairwise heuristic explains a greater share of choice in the lab and field than prior benchmarks.

Overview and Implementation Details. We administered the experiment (Experiment B) in July 2022 on the Qualtrics platform to 893 employed US adults, aged 25 to 65, recruited from Amazon Mechanical Turk. After describing the real-life GQ paradigm to participants, we randomized the 82 percent who successfully completed multiple comprehension checks to one of two experimental arms. Across both arms, participants were asked to make a hypothetical decision from a GQ menu populated with sales goals (105 units, 110 units, 115 units) and dollar-denominated rewards ($150, $450, $900) representative of (percent-denominated) programs from the field data.[24] To increase the realism of the hypothetical goal choice in the first arm, we provided participants a series of fictional sales figures for the prior 14 periods (in the second arm, we reported explicit likelihoods of attainment). The distribution of prior sales was engineered to produce an implied likelihood of goal attainment equal to that observed in the field. An exploratory pilot study led us to believe that participants would make hypothetical decisions in the context of vignettes in a manner resembling employees making real-life decisions in the field.

The first arm was designed to test whether participants adopted the pairwise heuristic in their evaluation of menu options and whether the degree of inferential bias predicted goal choice, conditioned on beliefs. Specifically, after participants selected their goal, we asked them to introspect as to how they arrived at their goal choice by asking them to indicate which, if any, pairwise comparisons they made during their deliberation (e.g., "At some point, I directly compared Goals 1 and 2"). We then asked participants to report beliefs of goal attainment through both a contingent and non-contingent elicitation. For the former, we asked participants to estimate, on a scale from 0 to 100 percent, their likelihood of Goal 3 attainment given certain knowledge they would attain Goal 2 and the analogous estimation for Goal 2 contingent on Goal 1 attainment. (Due to the hypothesized difficulty of credibly eliciting contingent expectations we piloted different communication strategies before arriving at the implementation used in the experiment.[25]) Finally, to generate between-subject evidence for the potential bias and to test its generalizability, we elicited contingent and non-contingent weather forecasts (across both experimental arms).[26] Across elicitations, validation rules disallowed internally inconsistent beliefs.

---

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

[25] 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?"

[26] We randomized participants to either forecast the likelihood that tomorrow's high temperature would be at least 70, 80, and 90 degrees Fahrenheit or to forecast the conditional likelihood of at least 90 degrees given certain knowledge of at least 80 degrees.

The second arm was designed to test whether the optimality of choice varied across menus intended to either emphasize or defuse the through the varying display of probabilistic information about goal attainment. Specifically, we randomized participants to one of three variations of the same representative menu. A first menu, *non-contingent display*, 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 accurate average attainment statistics from the field. A second menu, *contingent display*, 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 display*, 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, $\hat{s}_{h|l} = s_h$ (i.e., 89 percent and 88 percent from the menu with unbiased contingent likelihoods was replaced by 74 percent and 65 percent, respectively).

Results. The experiment yielded several pieces of evidence suggesting the use of the proposed heuristic by a substantial share of participants (baseline choice across the three goals: 0.24, 0.45, 0.31). First, we found significant evidence for the two 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 proximal comparison. More critically, participant estimates of contingent likelihoods revealed substantial and pervasive underestimation relative to the likelihoods implied by non-contingent estimates. For example, participants underestimated the contingent likelihood of attaining Goal 3 given attainment of Goal 2 by 22 percent relative to the likelihood implied by the non-contingent elicitation (0.59 relative to 0.76) and underestimated the likelihood of (Goal 2 | Goal 1) attainment by 23 percent (0.64 relative to 0.82). The between-subject estimates of weather implied even more severe underestimation, 38 percent, of contingent probabilities.

Second, we found that the magnitude of the bias in estimates of contingent goal attainment strongly predict optimal goal choice even after controlling for non-contingent beliefs of attainment. Specifically, we estimated a simple additively linear model of optimal goal choice, g_c^* , relative to predictions of a subjective EU risk-neutral benchmark, as a function of beliefs and inferential bias:

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

The parameter $\hat{s}_k$ 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 strongly predicted by the perceived likelihood of Goal 3 attainment ($b = 1.00$, $p < 0.001$) and the magnitude of the contingent bias in Goal 3 beliefs ($b = -0.81$, $p < 0.001$) (excluding six observations with a non-unique optimum). Given the overall share of optimal choice, 0.37, the estimates imply that eliminating the inferential bias, which averages 0.17 across the sample, would increase optimal choice by 37 percent (i.e., (-0.17 x -0.81) / 0.37). The considerable partial correlation between the contingent error and inference and optimal choice was evidenced across a variety of alternative non-parametric specifications.

Lastly, we document a marked increase in the optimality of participant choice from menus designed to diminish bias in contingent inference. Specifically, when engaging a menu with non-contingent display, participants selected the EV-maximizing goal at a rate of 41 percent. However, when selecting from a menu with an unbiased contingent display, 61 percent of participants selected the optimal goal, a 48 percent increase in optimal choice relative to the non-contingent display ($p = 0.002$). As further evidence for the importance of the inferential bias and the plausibility of its presumed magnitude, participants who engaged the menu displaying contingent likelihoods with bias made optimal choices at a rate, 39 percent, statistically indistinguishable from the non-contingent menu ($p = 0.82$) and the menu with no information display from the experimental first arm ($p = 0.55$). Beyond prompting a substantial increase in optimal choice, the debiased menu also reduced the choice of Goal 1, the lowest-EV option, by 27 percent (14.4 to 10.5) relative to the baseline menu with non-contingent display.

6.4. Descriptive Accuracy of Pairwise Heuristic – Lab and Field

Perhaps the most informative diagnosis of the pairwise heuristic is the accuracy with which it explains goal choice relative to other benchmarks. To assess the explanatory power of the heuristic in the field and lab, we must specify additional details as to how individuals might apply the heuristic to a GQ menu. First, in the field, unlike the lab where we can calculate a person-specific bias (because we directly observe both contingent and non-contingent beliefs of attainment), we must assume a particular functional form. Given the size of the bias exhibited by lab participants, we assume that employees in the field are subject to a bias of the form $\hat{s}_{h|l} = s_h$. Second, across lab and field, we specify noise allowances, $\varphi \in \{0, [-25, 25], [-50, 50]\}$, a range that spans plausible degrees of computational imprecision.[27] Next, as alluded to earlier, we assume decision-makers practically apply the heuristic to a GQ menu by first comparing Goals 1 and 2, and if Goal 2 is preferred to Goal 1, comparing Goals 2 and 3. Finally, to avoid mechanically inflating optimal choice shares due to benchmarks that generate non-unique predictions, we

[27] 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.

characterize choice for both the unrestricted sample of decisions and a restricted sample of decisions for which the heuristic produces a unique goal choice.

Table 7 summarizes the descriptive accuracy of the pairwise heuristic relative to the baseline subjective-EV benchmark under varying allowances for noise. In the lab, using personalized measures of contingent beliefs, the heuristic explained up to 57 to 61 percent of choice in the restricted and unrestricted samples, respectively. The estimates imply an increase in explanatory power of 49 to 65 percent relative to baseline, comparable to the increase implied by experimental response to the debiased and baseline menus. In the field, the pairwise heuristic, assuming a noise parameter of $25, increased explanatory power relative to the baseline benchmark by 26 to 46 percent across the restricted (0.50 to 0.63) and unrestricted samples (0.50 to 0.73). Explanatory power increases by up to 66 percent given a noise allowance of $50 but we caution that a noise allowance of that magnitude yielded unique predictions for only 42 (parametric bias) to 61 (no bias) percent of the sample. Comparing the descriptive accuracy of the heuristic across personalized and parameterized formulations of bias in the lab suggests that field estimates of explanatory power may modestly underestimate the efficacy of the heuristic had we been able to observe personalized bias among employees. Across the lab and field, the explanatory power of the pairwise heuristic exceeds prior benchmarks.[28]

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 the baseline 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 baseline benchmark and 30 percent of choices unexplained by the baseline benchmark. Inversely, the baseline benchmark explained 69 percent of choices also explained by the heuristic and only 8 percent of choices unexplained by the heuristic.

Notably, a significant share of choice does not adhere to the standard benchmark or the tested alternatives, including the proposed heuristic. While surely some decisions, in the lab and the field, likely reflect confusion, inattention, or an otherwise random choice process, we can make informed speculations

[28] Among the presumed mechanisms underlying the heuristic, the assumption of inferential bias appears more critical in delivering increased explanatory power than the assumption of noise. Our assumed stopping role—i.e., selecting Goal 1 if it survived the Goal 1 vs. Goal 2 pairwise comparison—does not seem materially important for characterizing choice given the similarity in descriptive accuracy between the benchmark and the heuristic without bias or noise.

as to the decision strategies responsible for at least some share of residual choice. For example, we speculate, based on lab evidence, that many individuals narrowly applied their preferred decision rule to a subset of the menu—e.g., they compared Goals 1 and 2 only, having ruled out Goal 3 for some other reason. Specifically, roughly one-quarter of participants in Experiment B who made choices unexplained by the heuristic reported only comparing goals 1 and 2—an indication fully consistent with their final goal choice. Applying the pairwise heuristic only to the first two goals explains all but one of these decisions. While we do not know why an individual might initially exclude certain options, narrowly engaging a menu seems reasonable in the context of larger ordered menus where evaluating all options might be overly effortful or engage undesired choices (e.g., an insurance consumer who knows they do not want a very high deductible). As another example, Experiment A, where we observe multiple choices per individual, suggests that a small share of individuals selected the lowest goal regardless of that goal's relative economic value. Assuming these responses were sincere, this could reflect a variety of decision motives such as an exclusive desire to maximize the likelihood of gain or to minimize risk.

7 APPLYING HEURISTIC TO OTHER RISKY-CHOICE SETTINGS

An advantage of GQ is that it offers an opportunity to investigate the motives for financial risk taking in a setting registering high in simplicity, transparency, variation in financial stakes, and compliance/participation. While employee reward programs are of independent economic interest given their popularity—as evidenced, in part, by the 40 percent of Fortune 500 firms that have adopted GQ itself—we speculate the proposed heuristic may help to understand financial risk-taking in other domains featuring choice from a menu of ordered options. Such domains could include contingent labor contracts, betting/gambling, consumer loyalty programs, portfolio allocation decisions, equity option premiums, and demand for deductible-based health, property, and vehicular insurance. We highlight two of these domains of potential applicability in greater detail.

Insurance Demand. A first domain of potential applicability of the heuristic is insurance demand. Specifically, in deductible-based (e.g., vehicular, health, home) insurance settings where coverage is often mandatory, researchers have found demand for a degree of coverage difficult to reconcile through preference-based risk aversion alone (Barseghyan et al. 2018). A difference between insurance and GQ with relevance for application of the heuristic is the presence of non-contingent costs, reflected in differences in plan premiums. The implication is that, in contrast to GQ, where an employee evaluating options via contingent pairwise comparisons can ignore the potential of attaining no goal (since failing to attain any goal yields no reward regardless of choice), in insurance, a decision-maker must compare the relative value of options across contingent states. As such, the contingency neglect assumed by the pairwise heuristic predicts that a decision-maker will both underestimate the relative likelihood of a large, versus small, claim and fail to sufficiently adjust for the baseline likelihood of any claim.

As illustration, consider a stylized setting where a consumer is required to purchase insurance for a new home (e.g., because of a lender mandate). Suppose that practically she must choose from a menu of two policies with either Low or High coverage that are identical but for their cost-sharing and annual cost (Low: \$1,000 deductible, \$600 premium; High: \$500 deductible, \$700 premium). Assume the expected likelihood of any (covered) damage to the home next year, $\hat{s}_{any}$, is 4 percent, comprised of a 3 percent chance of severe damage (\$1,000+), $\hat{s}_{severe}$, and a 1 percent of mild damage (\$250), $\hat{s}_{mild}$. Further assume both plans fully cover annual damage in excess of the deductible. The pairwise heuristic stipulates that a risk-neutral consumer would select the high-coverage plan by assessing, with some noise, φ, whether the expected potential benefit of increased coverage given severe damage under the focal contingency of some damage, $\Delta b_{h,l|claim}$, exceeds the plan's higher costs, $\Delta p_{h,l}$.

$$\hat{s}_{any}\left[\left(\hat{s}_{severe|any} * \Delta b_{h,l|severe}\right) - \left(\hat{s}_{mild|any} * \Delta b_{h,l|mild}\right) + \varphi\right] > \Delta p_{h,l}$$

With well-calibrated beliefs and the absence of noise, a decision-maker would, by some margin, reject the high-coverage plan (0.04[(0.75 x \$500) – (0.25 x \$0)] < \$700 - \$600). However, if afflicted by contingency neglect, and computational imprecision, the decision-maker would underestimate the relative likelihood that any damage will be severe and insufficiently adjust for the baseline likelihood of a claim. In the extreme, full contingency neglect simplifies the decision rule to: $\left(\hat{s}_{severe} * \Delta b_{h,l|severe}\right) + \varphi > \Delta p_{h,l}$. The simplified rule implies the choice of the high coverage plan so long as $\varphi > -275$.

To explore whether the pairwise heuristic can help to explain risk-averse (and excessively heterogeneous) deductible-based insurance choice, we administered a final experiment (Experiment C) to test whether menus encouraging non-contingent evaluation increased demand for less expensive low-coverage plans. We asked 435 US adults, aged 25 to 55 years, recruited from Amazon Mechanical Turk to select a hypothetical insurance plan for their new home from a menu of three actuarially overpriced options, adapted from Sydnor (2010). The (lender-mandated) plans varied 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 randomized participants to one of four menus that, while all featuring the same three plan options, varied the presence and framing of loss probabilities. The displayed probabilities, and subsequent evaluation of plan choice, assume figures roughly approximated from Sydnor (2010): a 4 percent overall likelihood of damage, comprised of a 3 percent likelihood of severe damage (\$2500+) and a 1 percent likelihood of non-severe damage. A first, *baseline menu*, featured no probabilistic information. A second, *focal contingency menu*, displayed the 75 percent conditional likelihood of severe damage, given any claim, and specified its costs, but did not display the base-rate likelihood of a claim. A third, *focal + non-focal contingency*, displayed both the 75 percent conditional likelihood of severe damage (and specified its costs) and the 4 percent claim base-rate. The fourth, *non-focal contingency menu*, only displayed the claim base-rate, this time as 96 percent, to heighten its salience. The second menu was intended to encourage relative evaluation within the focal contingency while the fourth menu was intended to encourage focus on the non-focal contingency. The third menu encouraged relative evaluation within the focal contingency while also communicating the base-rate.

As reported in Appendix Table A4, participants chose a diversity of plan options across menus, despite the low expected economic value of the medium and premium plans. Across the first three menus, more than one-half of participants chose something other than the least-expensive high-deductible plan. When faced with a menu encouraging engagement of the focal contingency (menus 2 and 3), participants were significantly less likely to select the least-expensive plan and far more likely to choose the most-expensive, low-deductible, plan, than a menu with no information display (menu 1) or a menu emphasizing the non-focal contingency (menu 4). The fourth menu led to the highest share of least-expensive plan choice, the EV-maximizing plan under our loss assumptions. Notably, across the last two menus—menus for which no set of beliefs can rationalize anything but basic plan choice for an EV-maximizing consumer—the menu emphasizing the non-focal contingency led to a 54 percent increase in optimal plan choice (from 0.35 to 0.54) and a 69 percent reduction in the most financially inefficient plan choice (from 0.26 to 0.08).[29] 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.

Portfolio Allocation. A second domain of potential applicability of the heuristic is portfolio allocation and participation in the stock market. Researchers have offered evidence that participation in the stock market is lower than predicted by most life-cycle models of consumption (see Gomes et al. 2020 for review) and that portfolio allocation decisions, particularly in retirement, are sensitive to the configuration of the choice menu (Benartzi and Thaler 2007). The pairwise heuristic offers a possible explanation for non-standard allocations from fund menus often faced by 401(k) enrollees who do not abide a default option. To illustrate, consider that menus of investment funds are typically ordered from lower to higher financial risk—e.g., inflation-protected bonds to small-cap equity funds, conservative funds to aggressive growth funds, sooner-dated target date funds to later-dated target date funds. In theory, while such decisions should be governed by factors including preferences for risk and

[29] While beliefs of improbably high base-rate likelihood of a claim could rationalize demand for more expensive plan options in the first two menus, in the last two menus, because the base-rate is specified, the most pessimistic, relative to the most optimistic, expectations of cost would only shift expectations about the 1 percent likelihood cost of non-severe damage (from $2,499 to $1).

expectations regarding future returns (e.g., the CAPM model of Sharpe, 1964; and Lintner, 1965), in the same manner that the pairwise heuristic led to excessive risk aversion in GQ goal choice (relative to standard benchmarks), its application could lead to overly conservative portfolio choice. Such conservatism would arise if investors, when deciding between funds with varying equity exposure, systematically underestimated the likelihood of very positive market returns, relative to positive returns (and further neglected evaluation assuming non-positive returns).

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 across potential contingencies. The heuristic posits that comparisons are subject to inferential error due to a neglect of non-focal contingencies. In the context of GQ, this neglect leads employees to underestimate high-goal attainment and 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 the possibility that household financial decisions in other menu-based domains, such as insurance or portfolio allocation, 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 engagement of non-focal contingencies differed from a baseline menu or a menu encouraging relative evaluation within the focal contingency. 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 underlying 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.

9 REFERENCES

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

Baillon, A., Bleichrodt, H., and V. Spinu. (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., and 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.

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

Benjamin, Daniel J. (2019). Chapter 2 - Errors in probabilistic reasoning and judgment biases. In B. Douglas Bernheim, Stefano DellaVigna, David Laibson (eds.), *Handbook of Behavioral Economics: Applications and Foundations 1, Volume 2* (pp. 69-186). North-Holland.

Bhargava, S., G. Loewenstein, 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.

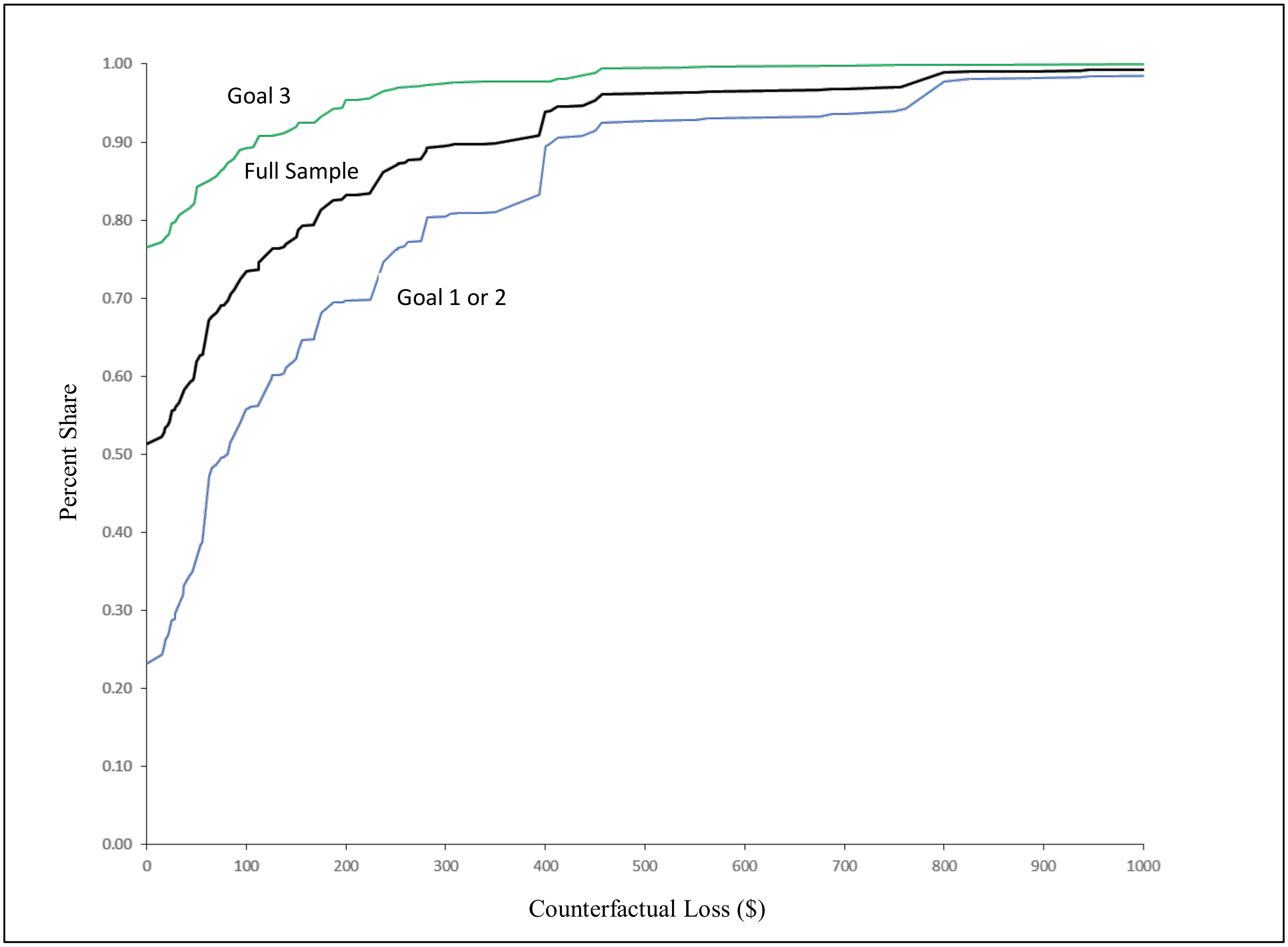
Bombardini, M., and F. Trebbi. (2012). Risk aversion and expected utility theory: An experiment with large and small stakes. *Journal of the European Economic Association*, 10(6), 1348–1399.

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.
Cohen, A. and L. Einav (2007). Estimating risk preferences from deductible choice. *American Economic Review*, 97(3), 745–788.
Chetty, R., and A. Szeidl. (2007). Consumption Commitments and Risk Preferences. *The Quarterly Journal of Economics*, 122(2), 831–877.
Cutler, D., and R. Zeckhauser. (2004). Extending the Theory to Meet the Practice of Insurance. In R. Litan and R. Herring (eds.), Brookings-Wharton Papers on Financial Service. Washington, D.C.: Brookings Institution.
Damasio, A. (2005). Descartes' Error: Emotion, Reason, and the Human Brain. New York, NY: Penguin.
Ericson, K., and A. Starc. (2012). Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts Connector. *The American Economic Review*, 102 (3): 493-97.
Gomes, F., M. Haliassos, and T. Ramadorai. (2021). Household Finance. *Journal of Economic Literature*, 59(3), 919-1000.
Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica*, 59(3), 667–686.
Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7), 2643–2682.
Handel, B. R. 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.
Hey, J., and C. Orme. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica*. 62, 1291-1326.
Holt, C., and S. K. Laury. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92 (5), 1644-1655.
Jaspersen, J. G., Ragin, M. A., and J.R. Sydnor. (2022). Predicting insurance demand from risk attitudes. *Journal of Risk and Insurance*, 89, 63–96.
Kahneman, D., and A. Tversky. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.
Kamenica, E. (2008). Contextual Inference in Markets: On the Informational Content of Product Lines. *The American Economic Review*, 98(5), 2127–2149.
Köbberling, V., and Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1), 119-131.
Koszegi, Botond, and M. Rabin. (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, 121 (4): 1133–65.
Koszegi, Botond, and M. Rabin. (2007). "Reference-Dependent Risk Attitudes." *The American Economic Review*, 97 (4): 1047–73.
Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review*, 90, 426–432.
Loewenstein, G., and J. Lerner. (2003). The role of affect in decision making, in Handbook of Affective

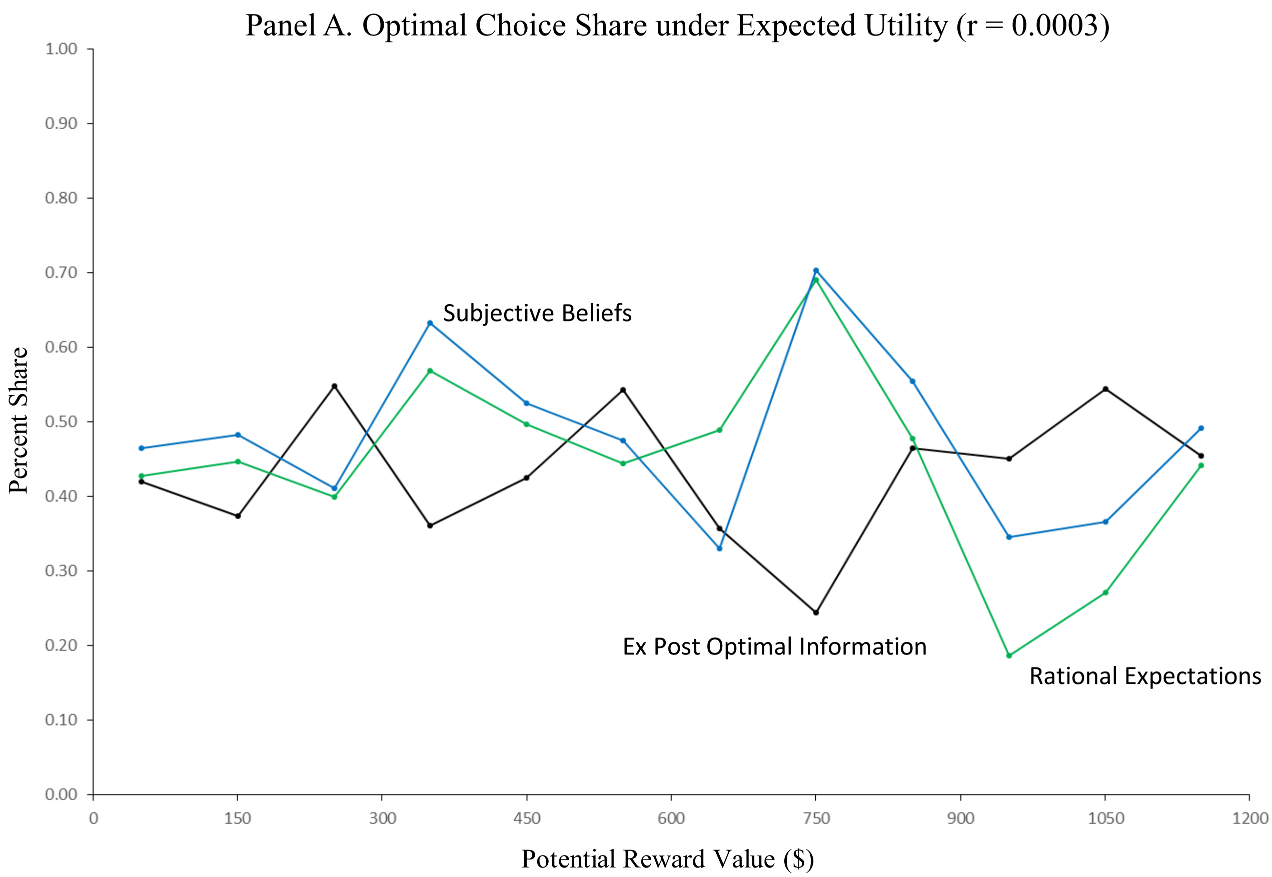
Science, eds R. Davidson, K. Sherer, and H. Goldsmith (Oxford: Oxford University Press), 619–642.
Loomes, G., and R. Sugden. (1982). Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal*, 92, 805–824.
Martínez-Marquina, A., Niederle, M., and E. Vespa. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109(10): 3437-74.
Niederle, M. and L. Vesterlund. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much?, *The Quarterly Journal of Economics*, 122(3): 1067-1101.
Post, T., Assem, M., Baltussen, G., and R.H. Thaler. (2008). Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review*, 98(1): 38–71.
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.
Sunstein, C.R., and R.J. Zeckhauser. (2010). “Dreadful possibilities, neglected probabilities.” *The Irrational Economist. Making Decisions in a Dangerous World*. Eds. Erwann Michel-Kerjan and Paul Slovic, Public Affairs Books, NY: New York.
Sydnor, J. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics*, 2 (4): 177–99.
Tversky, A., and D. Kahneman. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5 (4): 297–323.
Tversky, A., and D.J. Koehler. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*(4), 547–567.
von Neumann, J., and O. Morgenstern. (1947). Theory of games and economic behavior (2nd ed.). Princeton, NJ: Princeton University Press.

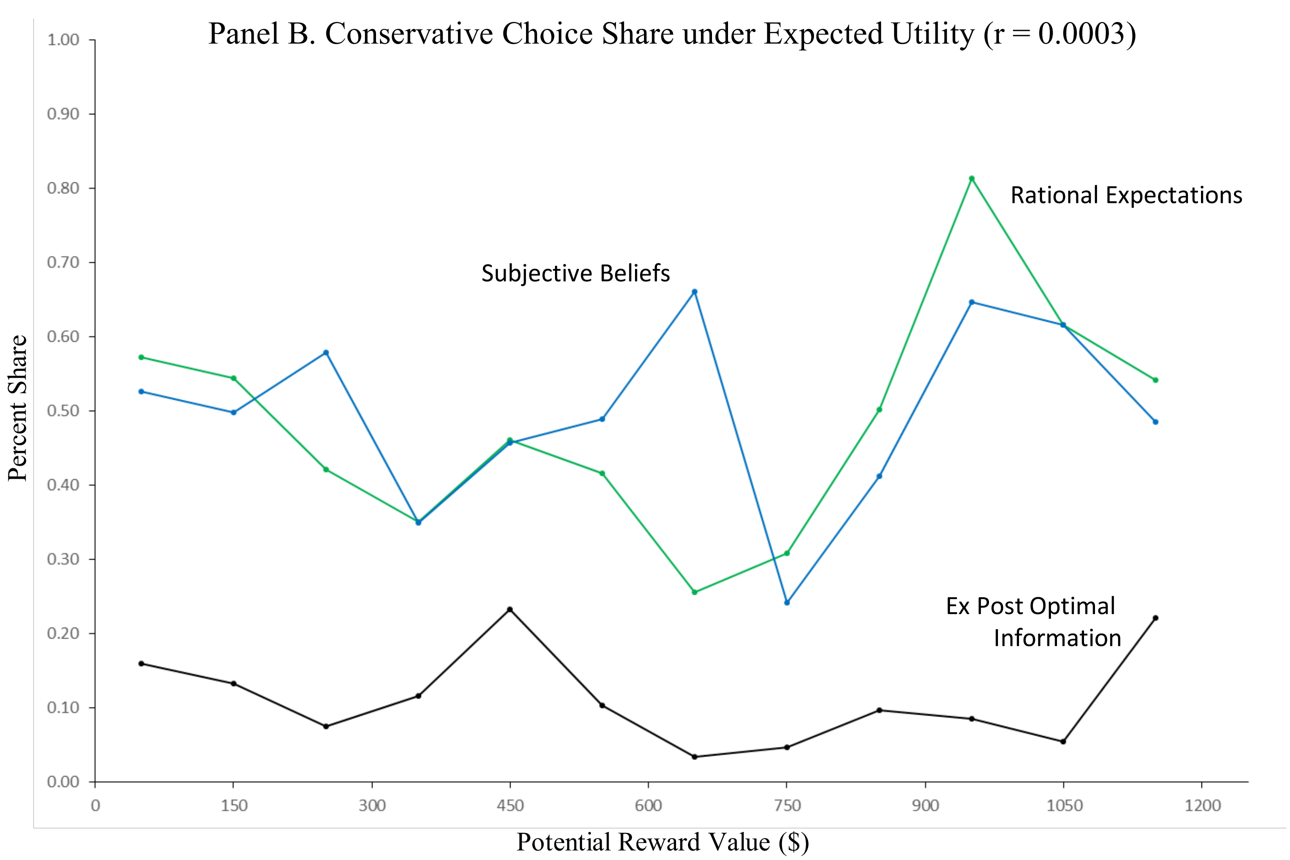
Figure 1.
Cumulative Distribution of Counterfactual Loss relative to Ex Post Optimal Choice | Goal Attainment



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

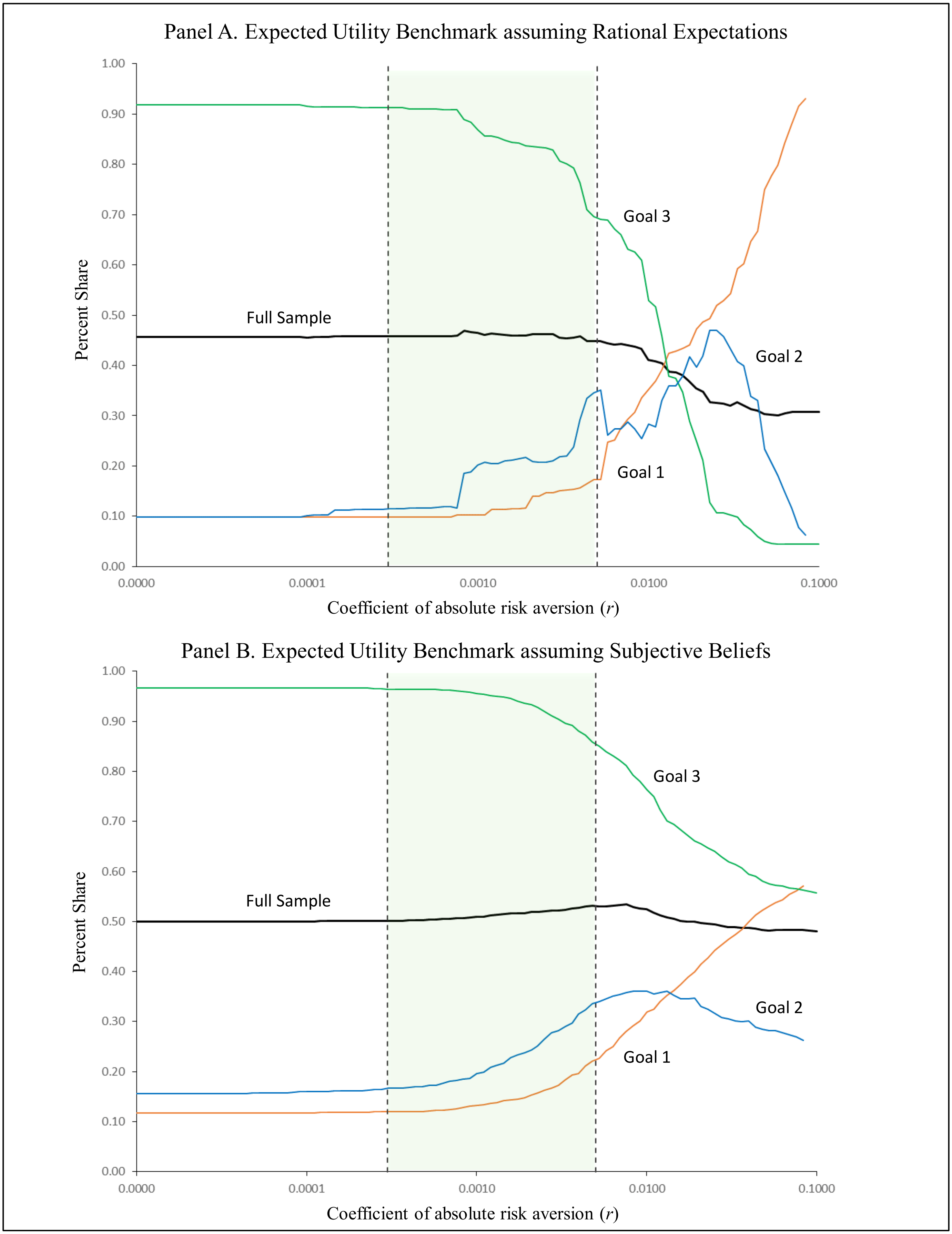
Figure 2.
Optimal and Conservative Choice under Expected Utility Benchmark by Potential Reward and Information Regime





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.

Figure 3.
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, $r \in [0.0003, 0.05]$.

Table 1.
Summary of Sample, Group and Employee Characteristics

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

Notes: This table summarizes observable detail on GQ programs and employees. 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

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

Notes: This table summarizes goal choice, productivity, and goal attainment for the 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

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

Notes: This table summarizes employee beliefs and confidence 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

	Risk Neutral EU		Expected Utility (CARA)					
			Rational Expectations			Subjective Beliefs		
	Rational	Subjective	r = 0.0003	r = 0.005	r [0, 0.005]	r= 0.0003	r = 0.005	r [0, 0.005]
Panel A. Characterization Overview								
Optimal Choice	0.45	0.50	0.45	0.44	0.56	0.50	0.53	0.59
Conservative Choice	0.49	0.48	0.49	0.38	--	0.48	0.39	--
Aggressive Choice	0.06	0.02	0.06	0.17	--	0.02	0.08	--
Panel B. Economic Consequences of Choice | Goal Attainment								
Counterfactual Loss								
Realized Reward	274	274	274	274	--	274	274	--
Counterfactual Reward | Ex Ante Optimal Choice	329	320	329	275	--	318	281	--
Counterfactual Loss (% of Counterfactual Reward)	0.17	0.14	0.17	0.00	--	0.14	0.02	--
Counterfactual Loss | Conservative Choice								
Realized Reward	164	162	162	122	--	159	118	--
Counterfactual Reward | Ex Ante Optimal Choice	303	281	302	244	--	276	222	--
Counterfactual Loss (% of Counterfactual Reward)	0.46	0.42	0.46	0.50	--	0.42	0.47	--
Panel B. Optimal Choice Share by Reward and Tenure								
Potential Reward Value								
Highest Quartile	0.42	0.48	0.42	0.39	--	0.49	0.55	--
Lowest Quartile	0.44	0.48	0.44	0.44	--	0.48	0.48	--
Employee Tenure								
Highest Category [10+ Years]	0.39	0.45	0.40	0.40	--	0.46	0.53	--
Lowest Category [< 1 Year]	0.44	0.47	0.44	0.44	--	0.47	0.50	--

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 (rational and subjective). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model. Panel B summarizes measures of the economic costs of choice conditioned on goal attainment. Panel C reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience. 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

	Subjective Expected Utility (CARA, r = 0.0003)			
	Baseline	Non-Linear Weights [Prelec, α = β = 0.65]	Noisy Beliefs [20% Error Band]	Composite Gain-Loss [RP = g; η = 1; λ = 2.25]
Panel A. Characterization Overview				
Optimal Choice	0.50	0.47	0.54	0.59
Conservative Choice	0.48	0.52	0.45	0.24
Aggressive Choice	0.02	0.01	0.01	0.17
Panel B. Economic Consequences of Choice | Goal Attainment				
Counterfactual Loss				
Realized Reward	274	274	274	274
Counterfactual Reward | Ex Ante Optimal Choice	318	324	317	272
Counterfactual Loss (% of Counterfactual Reward)	0.14	0.15	0.14	-0.01
Counterfactual Loss | Conservative Choice				
Realized Reward	159	163	150	168
Counterfactual Reward | Ex Ante Optimal Choice	276	282	269	296
Counterfactual Loss (% of Counterfactual Reward)	0.42	0.42	0.44	0.43
Panel C. Optimal Choice Share by Reward and Tenure				
Potential Reward Value				
Highest Quartile	0.49	0.44	0.55	0.61
Lowest Quartile	0.48	0.46	0.51	0.55
Employee Tenure				
Highest Category [10+ Years]	0.46	0.42	0.52	0.59
Lowest Category [< 1 Year]	0.47	0.44	0.51	0.59

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 the baseline SEU characterization (r = 0.0003). The next two columns characterize choice for the benchmark modified to allow for non-linear probability weights (Prelec 1998) and noisy beliefs, respectively. The final colum characterizes choice under the best-performing benchmark with gain-loss utility (see text for details). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model. Panel B summarizes measures of the economic costs of choice conditioned on goal attainment. Panel C reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience.

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

	Subjective Expected Utility (CARA, r = 0.0003)				Contextual Sorting Heuristics	
	Baseline	Non-Linear Weights [Prelec, $\alpha = \beta = 0.5$]	Noisy Beliefs [20% Error Band]	Composite Gain-Loss [RP = g; $\eta = 1$; $\lambda = 2.25$]	Ability	Taste for Competition
All Menus (6/6)	0.04	0.03	0.24	0.18	0.10	0.09
Nearly All Menus (5+/6)	0.16	0.13	0.42	0.40	0.12	0.10
All 3 Goal Menus (4/4)	0.16	0.09	0.29	0.25	0.13	0.12
All 4 Goal Menus (2/2)	0.10	0.10	0.44	0.40	0.15	0.15

Notes: This table characterizes the share of optimal 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 three columns characterize choice for a baseline subjective expected utility model assuming moderate risk aversion (r = 0.0003) as well as modified models incorporating non-linear decision weights and noisy beliefs. The next column features a composite gain-loss benchmark model (see text for details). A final set of columns characterizes choice for a set of heuristic-choice models involving contextual sorting by self-reported ability or taste for competition (see text for detail).

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

		Pairwise Heuristic - Bias in Contigent Inference								
	Risk Neutral	None			Personalized			Parameterized		
Decision Sample	SEU Baseline	[φ = \$0]	[φ = \$25]	[φ = \$50]	[φ = \$0]	[φ = \$25]	[φ = \$50]	[φ = \$0]	[φ = \$25]	[φ = \$50]
Field Data										
Restricted sample (unique predictions)	0.50	0.48	0.50	0.53	--	--	--	0.57	0.63	0.67
Unrestricted sample	0.50	0.51	0.58	0.63	--	--	--	0.57	0.73	0.83
Experiment B										
Restricted sample (unique predictions)	0.37	0.36	0.37	0.38	0.55	0.57	0.56	0.53	0.54	0.57
Unrestricted sample	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.

ONLINE APPENDIX

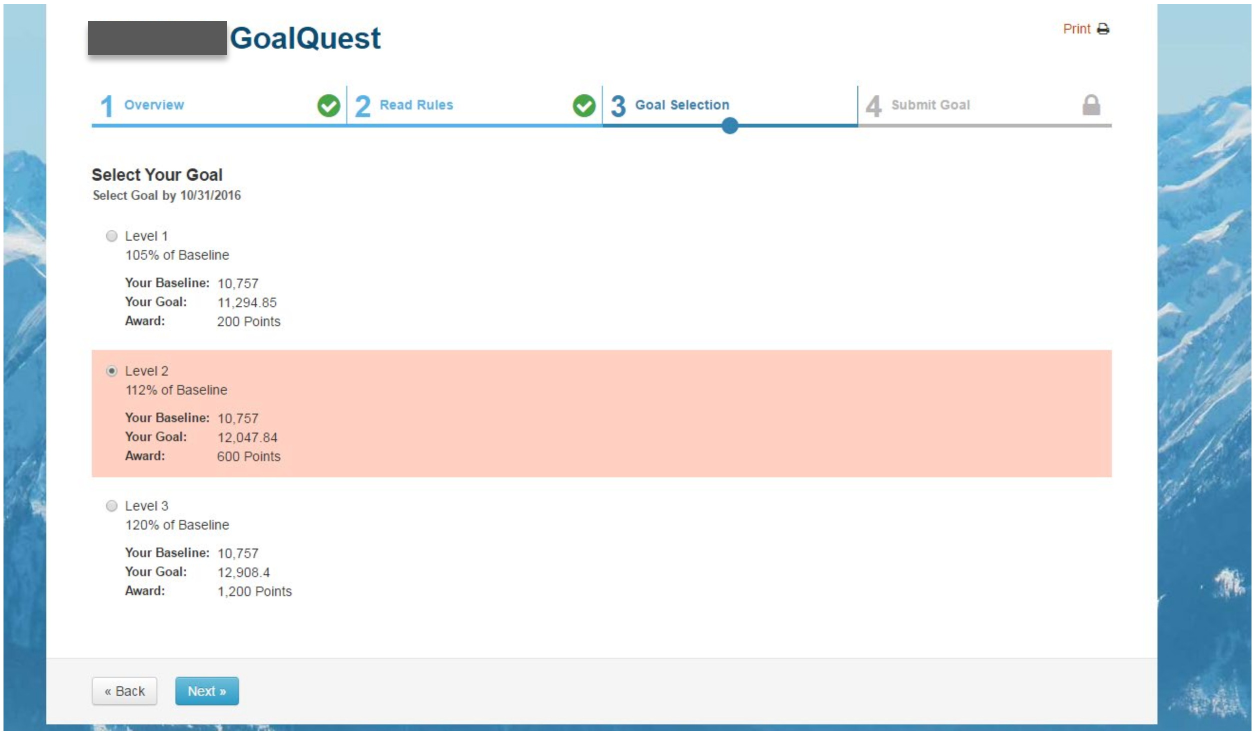
1.1 Characterizing Choice with CRRA Utility Benchmarks

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

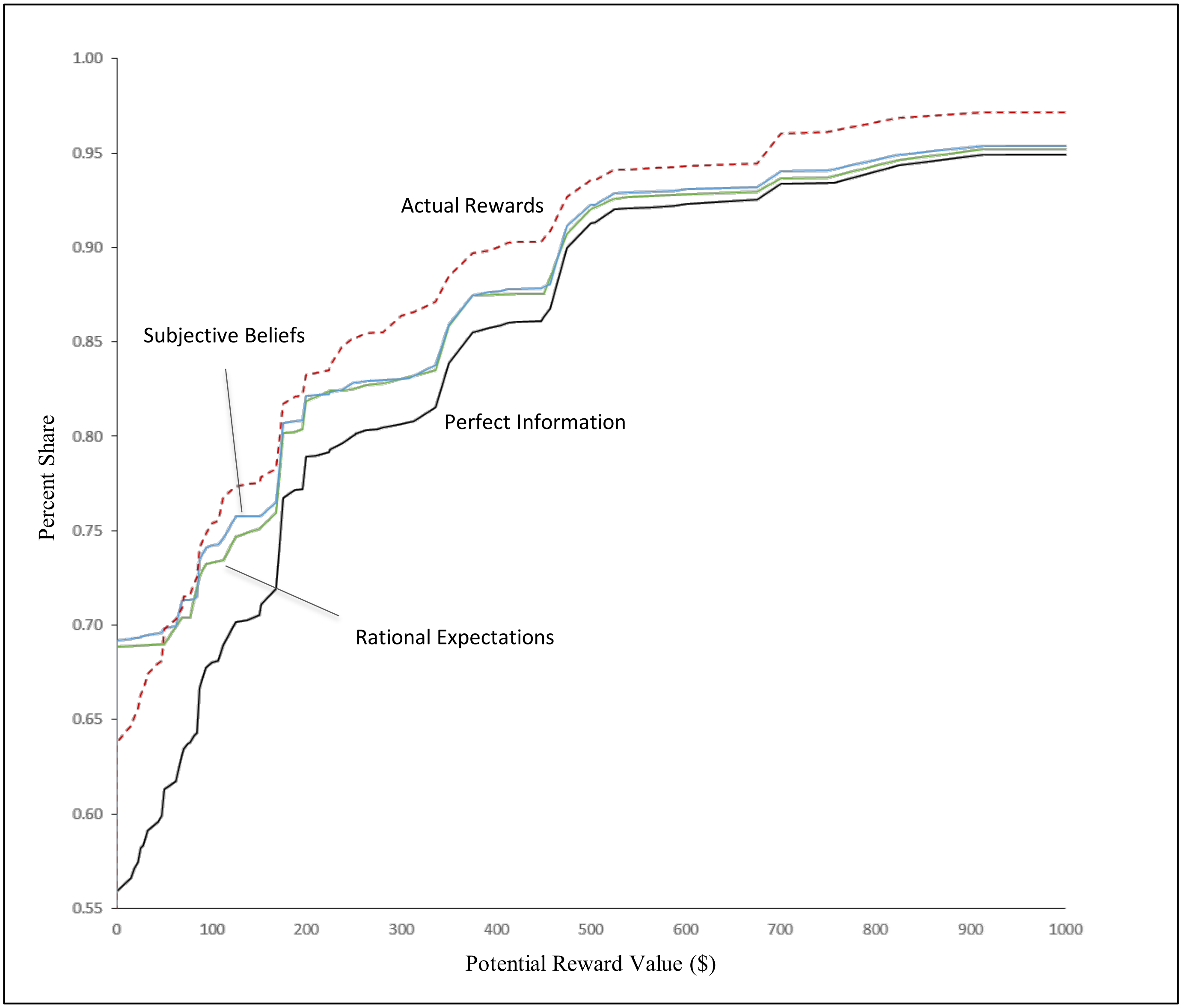
Specifically, we assume employees are governed by CRRA utility of the form: $u(x) = \frac{x^{1-\rho}}{(1-\rho)}$ for $\rho \neq 1$ and $u(x) = \ln(x)$ for $\rho=1$. We assess choice for initial lifetime wealth ranging from \$1,000 to \$1,000,000 and relative risk aversion, $\rho \in [0.10, 50]$. To appreciate the breadth of risk attitudes captured by the latter interval, we follow Post et al. 2007 in mapping risk parameters to the implied certainty coefficient—that is the certainty equivalent expressed as a fraction of expected value—associated with a 50/50 bet of (\$0, \$10k) assuming initial wealth of \$25,000. This interval almost certainly subsumes the range of plausible relative risk aversion—asserted by Holt and Laury (2002) as bounded by 0 and 1.37.

Appendix Table A2 summarizes the choice characterization by reporting the optimal choice share for the EU benchmark across beliefs (rational, subjective), initial wealth, and relative risk aversion. The table indicates that within the (highlighted) range of plausible attitudes towards risk (spanning certainty coefficients from 0.87 to 0.99), the CRRA benchmarks explain a share of choice virtually identical to the CARA analogues from Table 4 assuming either risk neutrality or moderate risk aversion for rational expectations (0.45) and subjective beliefs (0.50). Overall, we interpret the table as suggesting that one cannot attribute the lack of descriptive accuracy of the benchmark models tested in the main analyses to the assumption of constant absolute, rather than constant relative, risk aversion.

Appendix Figure A1.
Sample Image of GQ Goal Selection Interface



Appendix Figure A2.
Cumulative Distribution of Actual and Counterfactual Rewards under Risk Neutral Expected Utility Benchmark by Information Regime



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

Appendix Table A1.
Goal Choice Characterization for Gain-Loss Utility Benchmarks by Candidate Reference Point

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

Notes: This table 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.
Goal Choice Characterization for CRRA Utiity Benchmarks

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

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

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

Appendix Table A3.
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 A4.
Demand for Home Insurance across Information Displays - Experiment C

	Menu Display			
	Baseline	Focal Contingency	Focal + Non-Focal	Non-Focal Contingency
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 Cost [3% severe ($2500+), 1% non-severe ($500)]	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.