ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 452

Targeting Targets? Experimental Evidence on the Challenge of Individualizing Performance Targets

Saskia Opitz
Timo Vogelsang
Dirk Sliwka
Tom Zimmermann
Marina Talantceva

November 2026

www.econtribute.de

UNIVERSITÄT BONN
UNIVERSITÄT ZU KÖLN

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/2 – 390838866 – is gratefully acknowledged.

Targeting Targets? Experimental Evidence on the Challenge of Individualizing Performance Targets*

Saskia Opitz† Dirk Sliwka‡ Marina Talantceva § Timo Vogelsang¶ Tom Zimmermann‖

This Version: November 20, 2026

Abstract

This study examines whether tailoring performance targets to individual workers improves performance. We conducted an online experiment with nearly 3,000 participants, eliciting personality traits, preferences, and baseline performance in a real-effort task. One week later, they worked on the same task and were randomly assigned one of 38 target levels, receiving a fixed bonus upon reaching their target. Comparing performance across these target levels, we confirm a pronounced inverted U-shaped relationship between target difficulty and performance. We then apply machine-learning methods to estimate individually optimal targets based on participants' characteristics. Several characteristics interact with target difficulty, but the implied optimal targets cluster closely around the population optimum. Although predicted performance levels differ substantially across levels of baseline performance, the corresponding implied optimal targets vary little. Moreover, predicted performance functions are relatively flat around their peaks. Consequently, gains from personalized target assignment are economically negligible relative to the best uniform target.

*We thank John List, Matthias Mahlendorf, Kirill Ponomarev, and an anonymous reviewer of the Management Accounting Section Midyear Meeting (MAS) for their feedback and comments. Moreover, we thank participants of the Doctoral Colloquium of the Annual Conference for Management Accounting Research (ACMAR) 2025, Young Scholars Program Workshop of the ECONtribute Cluster of Excellence 2025, ERMAC Young Scholars (EYS) 2025, and the Management Accounting Section Midyear Meeting (MAS) 2026 for helpful comments. We gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/2– 390838866. Any errors and all opinions are our own. The survey and the experiment were approved by an IRB. We pre-registered the experiment on socialscienceregistry.org with the ID AEARCTR-0014235.

†WHU - Otto Beisheim School of Management. Institute of Management Accounting and Control. Email: saskia.opitz@whu.edu

‡University of Cologne. Faculty of Management, Economics and Social Sciences. Department of Corporate Development. Email: sliwka@wiso.uni-koeln.de

§University of Cologne. Faculty of Management, Economics and Social Sciences. Department of Corporate Development. Email: mtalantceva@wiso.uni-koeln.de

¶Frankfurt School of Finance & Management. Department of Accounting. Email: t.vogelsang@fs.de

‖University of Cologne. Faculty of Management, Economics and Social Sciences. Email: t.zimmermann@wiso.uni-koeln.de

Keywords: RANDOMIZED CONTROLLED TRIAL, HETEROGENEITY, TREATMENT EFFECTS, TARGETS, BEHAVIOR
JEL classification: J33, M10, M40, M50, M52

1 Introduction

Performance targets are among the most widely used instruments of management control, and one of the most consequential choices in their design is the choice of optimal target difficulty (Locke and Latham [2002], Arnold and Artz [2015], Baron et al. [2016], Matějka and Ray [2017], Matějka [2018], Gordon et al. [2019], Kim et al. [2023]). Targets that are too easy fail to stretch effort, while targets that are too ambitious discourage it, resulting in a relationship between target difficulty and performance that is commonly described as an inverted U (Locke and Latham [1990], Arnold and Artz [2015], Baron et al. [2016]).

However, the level of target difficulty at which performance is maximized may not be the same for everyone. Individuals vary widely in ability, personality, and preferences, and several of these characteristics have been shown to moderate how people respond to assigned goals (Locke et al. [1981], Barrick and Mount [1991], Erez and Judge [2001], Bandura [2013]).

Taken together, these two observations suggest an appealing conjecture: if performance-maximizing target levels vary across individuals, then tailoring target difficulty to the individual employee–holding the job and task environment fixed–should outperform assigning the same target level to all employees. This conjecture has gained practical relevance as organizations increasingly have data and algorithmic tools available to implement such tailoring, and recent evidence indicates that the individualized assignment of management practices can indeed generate substantial performance gains (Andreoni et al. [2023], Opitz et al. [2025], Lipman et al. [2026]).

In this paper we put the conjecture to a direct test. We conduct a pre-registered online experiment with nearly 3,000 participants in two parts. In the first part, we elicit a broad set of individual characteristics, including the Big-5 personality traits, risk tolerance, loss aversion, self-esteem, ambitiousness, and task-specific self-efficacy beliefs, and we measure baseline performance in a real-effort task without incentives. One week later, participants work on the same task again and are randomly assigned one of 38 target levels, with a fixed bonus paid upon reaching the assigned target. The real-effort task we implement has been well established in the literature (Berger and Pope [2011], DellaVigna and Pope [2018], De Quidt et al. [2018], Asad

et al. [2023], Dinhof et al. [2024], Opitz et al. [2025]). Crucially, prior research using this exact task has demonstrated that workers' effort provision is highly responsive to the specific type of incentive scheme employed (DellaVigna and Pope [2018]). Furthermore, recent experimental evidence applying machine learning algorithms to this task confirms substantial heterogeneity in worker responses, establishing that different incentive schemes work best for different individuals and that targeted assignment can significantly outperform a single uniform scheme (Opitz et al. [2025]). The fine-grained exogenous variation in target difficulty, combined with rich pre-treatment covariates measured before any target was known, thus should provide a strong basis for detecting systematic heterogeneity in performance-maximizing target difficulty.

Our first result confirms the typical relationship documented in the prior literature: Performance follows a clear inverted U-shape in target difficulty, rising steeply over low and moderate targets and declining once targets become very ambitious. By randomly assigning participants to 38 target levels, our design allows us to trace this relationship non-parametrically, rather than infer curvilinearity from a small number of target levels or functional-form assumptions, as most prior studies have done (Arnold and Artz [2015], Baron et al. [2016], Gordon et al. [2019]). The estimated peak lies at roughly 2,800 points, well above mean baseline performance of roughly 1,700 when no target is set.[1] Target choice has substantial aggregate consequences: a low target set at 500, for instance, yields about 30% less average output than the optimal target and a excessively high target of 3,900, for instance, leads to a 5% performance loss.

Our second and central result is that, among individuals performing the same task, worker types differ markedly in performance levels (the curve shifts vertically), but the target difficulty at which performance is maximized changes little. Employing key ideas from the causal machine learning literature, we estimate individual-level performance functions to derive the target that maximizes predicted performance for each combination of worker characteristics, including demographics, elicited personality traits and preferences. While some traits measurably change how performance responds to target difficulty, they hardly change *which* target maximizes predicted performance: individualized optima remain tightly concentrated around the population-

[1] Our parametric estimate of the uniform optimal target based on a quadratic specification is 2,811, the non-parametric estimate (i.e. simply the target value that descriptively yields the highest performance) is 2,700.

optimal level. When we evaluate the resulting assignment rules against the best uniform target, the predicted gains from personalization amount to approximately 0.4% of the average performance under the average best target of 2,800. Heterogeneity exists, but it leaves little scope for practically valuable personalization.

We suggest two reasons for this pattern, both of which plausibly extend beyond our setting. First, individual characteristics shift the *level* of the performance function far more than the *location of its peak*. A worker's prior performance, for instance, is a powerful predictor of performance: for a given target, second-round output is substantially higher for participants who performed better in the first round one week earlier. Yet the target level that maximizes performance is essentially unchanged by prior performance. In the framework of Shchetkina and Berman [2024], this pattern corresponds to high cross-treatment correlation: outcomes vary across individuals, but the ranking of targets is similar, so targeting has limited value. As they emphasize, heterogeneous treatment effects – the condition most empirical work checks for – are necessary but not sufficient for personalization to pay. What matters is whether individuals differ in *which* treatment is best for them, not merely in how strongly they respond overall. In our data, prior performance, the strongest predictor, for instance moves the entire curve upward while leaving the performance-maximizing target essentially unchanged. Second, the performance function is rather flat in the direct neighborhood of the optimum. Consequently, small deviations from a worker's "perfect" target barely reduce their output. This stands in instructive contrast to the tailored assignment of qualitatively different incentive schemes (where personalization consists of choosing from a discrete set of alternatives rather than a continuous variable), where personalization has been shown to yield sizeable gains in the same task and a similar subject pool (Opitz et al. [2025]). Choosing the type of incentive may therefore be a first-order decision, whereas fine-tuning the level of a chosen instrument appears to be second order.

Our study contributes to three strands of literature. First, we provide what we believe, to the best of our knowledge, to be the cleanest evidence to date on the shape of the target-performance relationship, based on fine-grained exogenous variation rather than a handful of treatment cells. Most prior studies on the relationship between targets and performance either examine a limited

number of target levels (Locke and Latham [2002], Baron et al. [2016]), or, despite exploiting richer variation in self-set targets, rely on endogenous target variation (Gordon et al. [2019]). Other studies infer curvilinearity from regression models that impose a specific functional form (Arnold and Artz [2015]). Second, we contribute to the rapidly growing literature on the individualized assignment of treatments and management practices (Murakami et al. [2022], Hitsch et al. [2024], Opitz et al. [2025]). Much of this literature documents settings in which targeting is beneficial, and only a smaller recent strand asks when it is not (Shchetkina and Berman [2024]). We provide experimental evidence on target setting – a canonical management control – and show that individualizing target difficulty may have very limited value. Third, we speak to the practice of target setting in organizations (Dekker et al. [2012], Kim et al. [2023], Matějka et al. [2024]) by examining whether optimal target difficulty varies sufficiently across workers to warrant individualized target assignment. Our results imply that the returns to choosing the common target level approximately right dwarf the returns to individual tailoring. For practitioners this is arguably good news, as uniform targets economize on data collection, avoid horizontal fairness concerns among comparable workers, and sidestep the governance questions that individualized algorithmic target setting would raise.

The remainder of the paper proceeds as follows. Section 2 reviews the theoretical background. Section 3 describes the experimental design. Section 4 presents the average effects and the personalization analysis. Section 5 concludes.

2 Theoretical Background

Target setting is a foundational concept in accounting, organizational behavior, economics, and psychology, emphasizing the importance of targets for motivation and performance (Erez and Zidon [1984], Miner [1984], Pinder [1984], Earley et al. [1989], Locke and Latham [1990], [2002], Marginson and Ogden [2005], Locke and Latham [2013], Gopalakrishnan et al. [2015], Gosnell et al. [2020]).

The psychological underpinnings of this instrument are most prominently formalized by Goal Setting Theory in organizational behavior (Locke and Latham [1990]). Locke and Latham [1990], for instance, demonstrate that "do-your-best" targets often result in lower performance than specific high target, and that too ambitious targets tend to lower performance, resulting in an inverted U-shaped relationship between target level and performance (Arnold and Artz [2015], Baron et al. [2016], Gordon et al. [2019]). Several studies have examined how to set performance targets to maximize efficiency, focusing on both specificity and difficulty (Winters and Latham [1996], Dekker et al. [2012], Locke and Latham [2015], Matějka and Ray [2017], Matějka [2018], Kim et al. [2023]). These studies aim to identify the conditions under which performance is most effectively enhanced while minimizing the potential downsides of targets that are either too challenging or too easy. The general conclusion from this body of work is that optimal targets require both specificity to guide effort and a calibrated level of difficulty – challenging enough to motivate, but not so extreme as to cause performance to drop.

Recent research suggests that organizations can improve the effectiveness of management controls by tailoring them to predictable differences across employees (Becker-Peth et al. [2013], Andreoni et al. [2023], Opitz et al. [2025], Lipman et al. [2026]). Opitz et al. [2025], for instance, show that algorithmically assigning incentive schemes based on worker characteristics can outperform the best uniform scheme. This raises the question of whether similar gains can be achieved by individualizing target difficulty. Such gains require not only that individuals differ in the target levels that maximize their performance, but also that these differences can be predicted from observable characteristics.

More generally, heterogeneity is actionable for target assignment only if two conditions hold. First, observable worker characteristics must predict meaningful differences in the target level that maximizes performance, rather than merely differences in performance at a given target. Second, deviations from these individualized optima must be sufficiently costly. When performance functions are flat around their peaks, even correctly predicted differences in optimal targets generate little value. These conditions distinguish heterogeneity in outcomes from heterogeneity that warrants different managerial actions.

Guided by the literature, we identify a broad set of observable individual characteristics that theory and prior evidence suggest may shape how performance targets motivate employees. For one, higher targets tend to lead to higher performance levels when individuals possess higher levels of task-relevant skills/ability (Locke and Latham [1990], Matějka et al. [2024]) as commonly measured by prior performance on the task. Personal characteristics such as self-esteem (Yukl and Latham [1978], Locke et al. [1981]), self-efficacy (Latham et al. [1994], Bandura [2013]), and risk aversion (Knight et al. [2001], Larrick et al. [2009]) can also be key moderators of target-performance interdependence.

For instance, research indicates that individuals with higher self-esteem tend to exert greater effort without immediate rewards and achieve assigned targets more effectively after receiving feedback (Schrauger and Rosenberg [1970], Dossett et al. [1979], Latham and Saari [1979]). However, its influence on target setting remains debated, with some studies suggesting a positive link to ambitious targets (Erez [1977], Pilegge and Holtz [1997]), while others report mixed findings that appear to depend on other personal characteristics such as gender (Levy and Baumgardner [1991]). Beliefs about one's capability to perform the specific task further shape goal acceptance and persistence. Such self-efficacy is typically associated with greater commitment to difficult goals, implying a higher optimal difficulty (Stajkovic and Luthans [1998], Locke and Latham [2002]).

Concerning the Big-5 personality traits, a first mechanism relates to traits that plausibly support sustained effort. Conscientiousness – encompassing planning, diligence, and impulse control – suggests a tendency to persist longer under demanding goals, investing effort rather than disengaging quickly (Barrick and Mount [1991]). Such persistence implies that individuals high in conscientiousness may achieve relatively better outcomes even at higher target levels. Extraversion may operate in a similar fashion when targets are salient and progress is visible: outward-facing benchmarks appear to energize extraverted individuals (Judge and Zapata [2015]), plausibly shifting the performance optimum toward more demanding targets.

Attitudes toward uncertainty also provide relevant patterns. Distinguishing risk tolerance from loss aversion, greater tolerance of risk makes aggressive effort strategies more acceptable and can sustain motivation at higher levels of difficulty. By contrast, stronger loss aversion increases the disutility of falling short and may prompt earlier disengagement once goals appear unattainable (Heath et al. [1999], Larrick et al. [2009]). Together, these tendencies imply higher optimal targets under risk tolerance and lower optima under loss aversion.

Other traits plausibly dampen the benefits of stretch targets. Neuroticism, reflecting emotional instability, may crowd out effort as failure risk becomes salient, while agreeableness may reduce motivation from competitive or strongly evaluative benchmarks (Judge and Ilies [2002], Judge and Zapata [2015]). Both patterns suggest lower optimal difficulty on average.

Taken together, prior research identifies multiple observable characteristics that may predict heterogeneity in responses to target difficulty. It does not, however, establish whether these characteristics generate sufficiently large and predictable differences in performance-maximizing targets or whether deviations from those individualized targets entail meaningful performance losses. Our empirical analysis therefore examines both whether observable characteristics predict variation in optimal target difficulty and whether using that variation to personalize targets improves performance relative to a well-calibrated common target. Because characteristics with limited predictive power in isolation may nevertheless be informative when considered jointly, we use machine-learning methods to assess the full set of characteristics in a data-driven manner.

3 The Experimental Design and Procedure

The experiment consists of two parts. In the first part, participants complete a survey measuring individual characteristics and then work on the real-effort task for 10 minutes.

The survey elicited demographic information as well as the psychological traits and economic preference measures described above. The Big Five personality traits were measured using the Mini-IPIP scale (Donnellan et al. [2006]). We used standard validated items/scales to elicit risk tolerance (Dohmen et al. [2011]), self-esteem (Rosenberg [2015]), loss aversion (Li et al.

[2021]), and ambitiousness (Hirschi and Spurk [2021]). In addition, following Danz et al. [2022], we elicited participants' beliefs about their expected performance on the task. We interpret this measure as task-specific self-efficacy, that is, participants' perceived ability to perform successfully on the task.[2]

The real-effort task is a standard task used in experimental economics to measure subjects' effort (Berger and Pope [2011], DellaVigna and Pope [2018], De Quidt et al. [2018], Asad et al. [2023], Dinhof et al. [2024], Opitz et al. [2025]).[3] Participants have to alternately press "a" and "b" on their keyboards, and receive a point for each alternate pressing. Before the task begins, participants complete a 30-second practice round to familiarize themselves with the procedure. They then work on the task for 10 minutes, being asked to score as many points as possible but without a performance target or performance-based incentives. Prior research has demonstrated that workers' effort provision in this task is highly responsive to the specific type of incentive scheme employed (DellaVigna and Pope [2018]), that there is substantial heterogeneity in worker responses to different incentive schemes and that personalized assignment of incentive schemes can significantly outperform a single uniform scheme (Opitz et al. [2025]).[4]

Throughout the real-effort task, participants see the remaining time and their cumulative score. Performance in this first part serves as our baseline measure of task performance and is subsequently used in the analysis.

[2]In economics, such measures are typically interpreted as subjective expectations about task performance. In psychology, closely related constructs are discussed under the label of task self-efficacy (Bandura [1977]). Although the two literatures differ in emphasis, both approaches capture individuals' perceived ability to succeed in a specific task. The experimental screens and exact survey questions are provided in Appendix B.

[3]A very similar task has also been used in experimental psychology (Amir and Ariely [2008]).

[4]In this experiment, a machine learning algorithm assigned workers to one of three optimal incentive schemes: a bonus loss scheme, a competitive real-time rank feedback condition, or a social pay-for-performance scheme. This targeted approach led to a 22.5% larger treatment effect than the single best uniform scheme (the bonus loss scheme).

Before accepting the offer to take part in the first part of the study, participants had already been informed that the study consisted of two parts and that participation in the first part required an intention to complete both. After completing the first part, participants were again reminded that the study included a second part and that they were expected to participate in it as well. When the second part began a week later, participants were invited via email.[5] The experiment timeline is also presented in Figure 1.

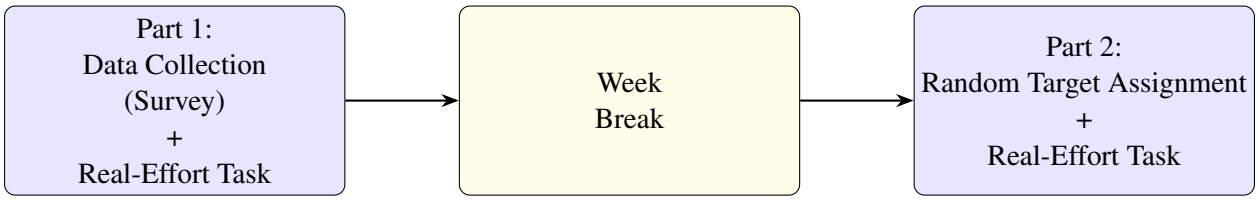

Figure 1: Timeline of the Experiment

Before starting the real-effort task in the second part, participants were randomly assigned a target specifying the number of points they were expected to reach. Target levels ranged from 200 to 3,900 points in increments of 100 points. Reaching this target was linked to a bonus of £0.80 amounting to approximately 50% of the fixed payment. After the end of the task, participants received information on their payment (fixed participation fee in the first part and fixed participation fee plus bonus for reaching the target in the second part respectively).

The assignment procedure was designed to ensure that most participants faced targets that were challenging yet achievable, rather than over-sampling extreme values that were either universally easy or virtually impossible. To accomplish this, targets were calibrated using baseline performance data from a pilot study and from Opitz et al. [2025]. These past scores were grouped into quintiles. Each participant was first randomly assigned to a quintile and then received a target randomly drawn from that specific range. This stratified approach naturally concentrated the target assignments in the most realistic performance ranges.

[5]We introduced this one week break for two reasons: i) we did not want participants' answers to the survey to influence consciously or unconsciously their performance in response to the assigned target, ii) as shown in Opitz et al. [2025], participants who interact again with the employers provide more consistent answers than one-shot participants.

We ran the experiment on the platform Prolific. The experiment took place between August 2024 and January 2025. Participation was restricted to US residents who speak fluent English. 3,702 participants were invited to the first part, and 3,095 out of those were invited to the second part [6]. In total, 2,804 participants completed the experiment with results[7]. 53% of those participants were male, 64% were college graduates, with an average age of 44. A graphical representation of the demographic factors can be found in Figure 5 in Appendix A.

4 Results

4.1 Average Effect of Target Difficulty

To assess how target difficulty affects performance on average, we first investigate the descriptive relationship between target difficulty and performance in the second part of our study. Figure 2 provides binned scatter plots and shows an inverted U-shaped relationship between the target level and the resulting performance in the second part of the experiment, in line with typical results on goal setting. The left panel reports average performance for each target level. Average performance increases from fewer than 1,400 points for targets below 500 points to a peak of approximately 2,000 points at a target of 2,800 points, before declining to about 1,900 points at a target of 3,900 points. Relative to the peak, this corresponds to a performance loss of more than 30% for very low targets and approximately 5% for very high targets. This pattern suggests that performance is highest when targets are challenging but remain attainable. The differences are substantial, with average performance varying by several hundred points across target levels. [8]

[6] Only the participants who have successfully completed the first part of the experiment were invited to the second part.

[7] Those are the participants who gained more than 0 and less than 4,000 points in each of the parts (if the participant reaches more than 4,000 points, they are considered to be bots). The average completion times were 24.2 minutes (median: 21.0 minutes) for the first part, and 12.8 minutes (median: 11.3 minutes) for the second part, respectively. Participants received a fixed payment of £2.40 for the first part and earned an average of £2.15 for the second part, resulting in compensation that falls within the typical payment range on Prolific.

[8] Average performance was 1,706 points in the part of the study without targets and 1,913 points in the part with randomly assigned targets.

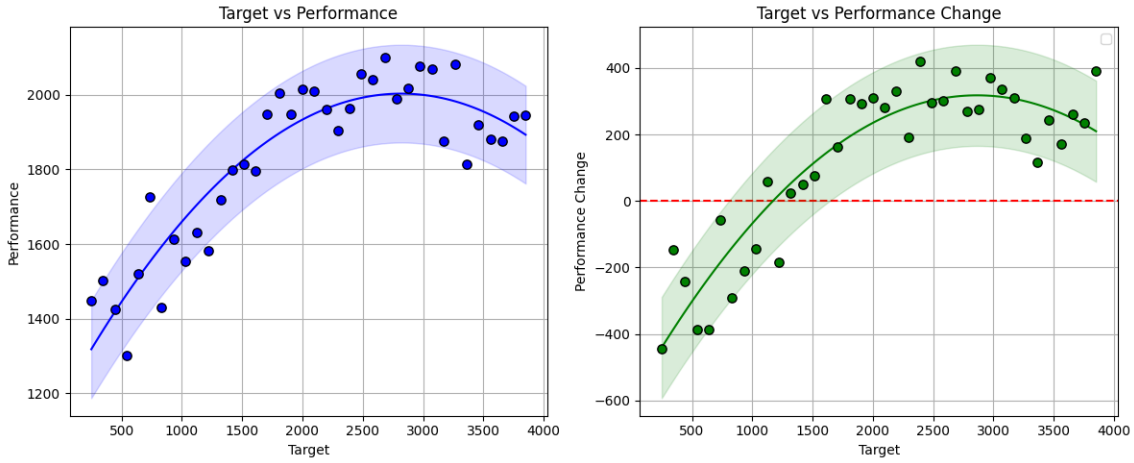

Figure 2: Average Performance and Performance Difference for Each Target Value

Note: This figure shows the relationship between target levels and performance. Each dot in the binned scatter plots represents the average performance (left panel) or average performance difference (right panel) per target level. The left panel shows the average performance under each target level in the second part of the study, while the right panel shows the difference between the performance in the first and the second part.

The right panel in Figure 2 illustrates the relationship between the target and the difference in performances between the first and second parts of the experiment. It therefore compares performance without a target or monetary incentive in the first part with performance under a randomly assigned target in the second part, demonstrating that participants with low targets, namely targets of 1,000 points or below, show, on average, a performance in the second part that is lower than in the first part where there had not been any target.

To assess this relationship more formally, we regress performance of a subject i on the respective $Target_i$ and its squared term $Target_i^2$ estimating the following model:

$$\text{Performance}_i = \beta_0 + \beta_1 \cdot \text{Target}_i + \beta_2 \cdot \text{Target}_i^2 + \gamma' X_i + \varepsilon_i \tag{1}$$

Table 1 shows the results controlling for prior performance only (column (1)) or further controlling for demographic characteristics and personality traits (column (2)). Both specifications provide further support for the inverted U-shaped relationship: the coefficients for $Target_i$ and $Target_i^2$ are highly significant. The estimated function attains a maximum at $T^\star = -\frac{\beta_1}{2\beta_2} = 2811$.

Table 1: Effects of Target on Performance

	Points Second Part			
	(1)	(2)	(3)	(4)
	All Observations	All Observations	Target below the median	Target above the median
Target	0.622***	0.633***	0.295***	-0.104**
	(0.052)	(0.052)	(0.021)	(0.047)
$Target^2$	-0.000***	-0.000***		
	(0.000)	(0.000)		
Points First Part	0.598***	0.597***	0.555***	0.651***
	(0.018)	(0.019)	(0.022)	(0.029)
Controls	-	x	-	-
Observations	2804	2804	1540	1264
R^2	0.356	0.368	0.375	0.326

Note: In this table, we report the results of regressions of Performance (points gained in the second part of the experiment) on Target and Prior Performance (points gained in the first part of the experiment). Column (1) shows results for the full sample, and column (2) includes controls which are demographic characteristics and personality traits. In column (3) and column (4), the sample is restricted to participants who got the target below (above) the median target level. Significance levels: $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$. Format of coefficient cell: Coefficient (Std. Error). Standard errors are heteroskedasticity robust.

As a second test, we perform a median sample split to investigate the relationship between target difficulty and performance in both subsamples separately. The results are presented in Table 1, columns (3) and (4). For the subsample of participants with a target below the median (2,600 points or below), we see a positive effect of the target difficulty on performance. For the subsample with targets above the median, we, however, see a negative association. Thus, our evidence clearly confirms the inverted U-shape suggested by the target-setting literature (Locke and Latham [2002], Arnold and Artz [2015], Baron et al. [2016], Matějka and Ray [2017]).

4.2 Tailoring Targets

4.2.1 Predictive Framework and Individual Optima

As a first step in examining whether heterogeneity in responses to assigned targets can be systematically predicted, we expand equation (1) to include interaction terms between targets (and their square) and personality traits. Results are reported in Appendix Tables 7 and 8.[9] The interaction terms are only statistically significant for Education, Loss Aversion, and Age. Thus, we find some statistical evidence of heterogeneity in responses to assigned targets, but this heterogeneity appears to be limited to a small number of individual characteristics.

Statistical significance, however, is not sufficient for heterogeneity to be actionable. An interaction may be precisely estimated yet shift the performance-maximizing target only slightly. Moreover, even meaningful differences in predicted optimal targets create little value when performance functions are flat around their peaks. The relevant question is therefore not only whether target responses vary across workers, but whether observable heterogeneity changes the optimal assignment enough to improve performance relative to the best uniform target.

Importantly, because characteristics with limited predictive power in isolation may nevertheless be informative when considered jointly, we next assess the predictive value of the full set of observable individual characteristics in more flexible models.

We evaluate and compare different machine learning models to predict second-part performance based on prior performance, personality traits, preferences and the target. In particular, we consider standard OLS regressions, Post-Lasso Regressions[10], and tree-based ensemble methods.

[9] To account for multiple hypothesis testing, we apply the bootstrap step-down procedure from List et al. [2019] and List et al. [2023] globally across all models to control the family-wise error rate (FWER). The adjusted results are reported in Appendix Tables 9 and 10.

[10] While standard OLS uses all included variables and risks overfitting, the Lasso algorithm prevents this by shrinking the coefficients of less predictive variables to zero, effectively dropping them. However, this shrinkage also biases the coefficients of the variables it keeps. Post-Lasso corrects this by using Lasso solely as a variable-selection tool in the first step, and then estimating a standard, unpenalized OLS regression on the selected variables in the second step (Belloni and Chernozhukov [2013]).

Note that tree-based ensemble methods, such as Random Forests and XGBoost, are highly flexible non-parametric models capable of endogenously capturing complex interactions and non-monotonic patterns – hence they substantially relax functional form assumptions. To also enable linear models, such as OLS and Post-Lasso, to capture non-monotonic relationships, we expand equation (1) to include interaction terms between targets (and their square) and personality traits.

Each model is trained and evaluated on 100 different random train-test splits, with 25% of the data in the respective test sample. The reported values summarize predictive performance across all out-of-sample test instances pooled over these 100 repetitions. Table 2 summarizes the predictive performance of the different models.

Table 2: R^2 and MSE Comparisons Across Models

	R^2	MSE
Post-Lasso (No Interactions)	0.3586	340,321
OLS (No Interactions)	0.3580	340,626
Post-Lasso (With Interactions)	0.3538	342,327
OLS (With Interactions)	0.3500	344,377
Random Forest	0.3404	350,003
XGBoost	0.3401	350,167

Note: This table reports overall out-of-sample performance for each model, measured by R^2 and MSE, aggregated across all test observations. Each model is trained and evaluated on 100 different random train-test splits. Test sample size was 25% of the data. The reported values summarize predictive performance across all out-of-sample test instances pooled over these 100 repetitions. Interactions in OLS and Post-Lasso are interactions between target variables (target and target squared with personal characteristics).

The R^2 values for the first four specifications (two OLS and the two Lasso variants) are very similar and exhibit minor sensitivity to the choice of random seed. Depending on the specific draw, any of these models may slightly outperform the others. A first key result of this analysis is thus that including interaction terms does not consistently improve predictive performance, indicating that the simpler additive specifications (which thus impose the restriction that the estimated curves are only shifted vertically by the elicited traits) perform comparably to the more

complex ones that allow for the curvature and peak of the performance-target relationship to vary with worker characteristics. This already suggests that heterogeneous responses to assigned targets, as captured by interactions between target variables and individual characteristics, may contribute little to the prediction of performance.

For the remainder of the analysis, we (mostly) focus on the Post-Lasso model with interaction terms. Among the models that allow the effect of target variables on performance to vary with individual characteristics, this specification achieves the best predictive performance out-of-sample. It is thus particularly well suited to examine heterogeneous responses to assigned targets. We also repeat the analysis using a Random Forest model, which provides a more flexible specification by allowing for nonquadratic functional forms and rich interactions among individual characteristics and obtain very similar results.[11]

4.2.2 Post-Lasso model with Interaction Terms

To examine which individual characteristics, if any, are systematically associated with heterogeneous responses to assigned targets, we analyze the variables selected by the Post-Lasso model, with particular attention to the interaction terms. To understand the importance of the interaction terms, consider the following simple model. When performance follows a standard quadratic function, $y = \beta_1 t + \beta_2 t^2 + \beta_3 x + \beta_4 (t \times x)$ for target t and characteristic x, the performance-maximizing target is given by $t^* = \frac{-\beta_1 - \beta_4 x}{2\beta_2}$. Hence, an individual characteristic x shifts the location of the optimal goal if and only if its interaction with the target variable is non-zero ($\beta_4 \neq 0$); otherwise, the trait merely shifts the curve vertically but does not move the target that maximizes performance.

Figure 3 reports how frequently each variable is selected by the Lasso procedure across 100 estimations with different random sample splits into train and test data. The variables selected in every specification are agreeableness, extraversion, prior performance – measured by points earned in the first part of the experiment – the target and its squared term, and the interaction between the target and prior performance. In addition, the interactions of the squared target

[11] Results are reported in Figures 7, 8, and 9 and Tables 5 and 6 in Appendix A.

with agreeableness and college graduation, as well as the interaction between the target and ambitiousness, are selected in more than 80% of the train–test splits. These selection frequencies suggest that at least some individual characteristics are systematically associated with how subjects respond to different target levels. [12]

Individualized optimal targets To better understand how personality-target interactions affect predicted performance and which target level maximizes predicted performance, we use the results of the Post-Lasso model to derive individually optimal targets based on participants' observable characteristics.

The procedure consists of two steps. First, for each individual, we predict performance at every target level and average these predictions across model estimations based on different train-test sample splits (see Section 4.2). This yields an individual-specific target–response curve. Second, we identify the target level that maximizes predicted performance. Formally, the optimal target is defined as the value that yields the highest predicted outcome conditional on the individual's characteristics:

$$g^*(X) = \arg\max_{g} \ \hat{y}(g, X).$$

Using these individual-specific optimal targets, we then quantify the predicted performance loss from assigning all individuals the common best target of 2,800 points rather than their individually optimal target.[13] The average predicted performance loss in our sample is 8.37 points, with a standard deviation of 14.31 points. Relative to the average performance of 1,913 under randomly assigned targets, this corresponds only to approximately 0.44%, and relative to the average performance of 2,028 under the best average target, this corresponds only to approximately 0.41%. Thus, while heterogeneity seems to exist, it does not result in sizable predicted performance differences between the individual best targets and the common best target.

[12] Our findings are consistent with prior research on target difficulty, which stresses the role of prior performance (i.e., the ability to perform a given task) (Locke and Latham [2002], [2013]) and provides some evidence for the role of extraversion (Judge and Zapata [2015]).

[13] Figure 10 in the Appendix reports the resulting differences for all observations.

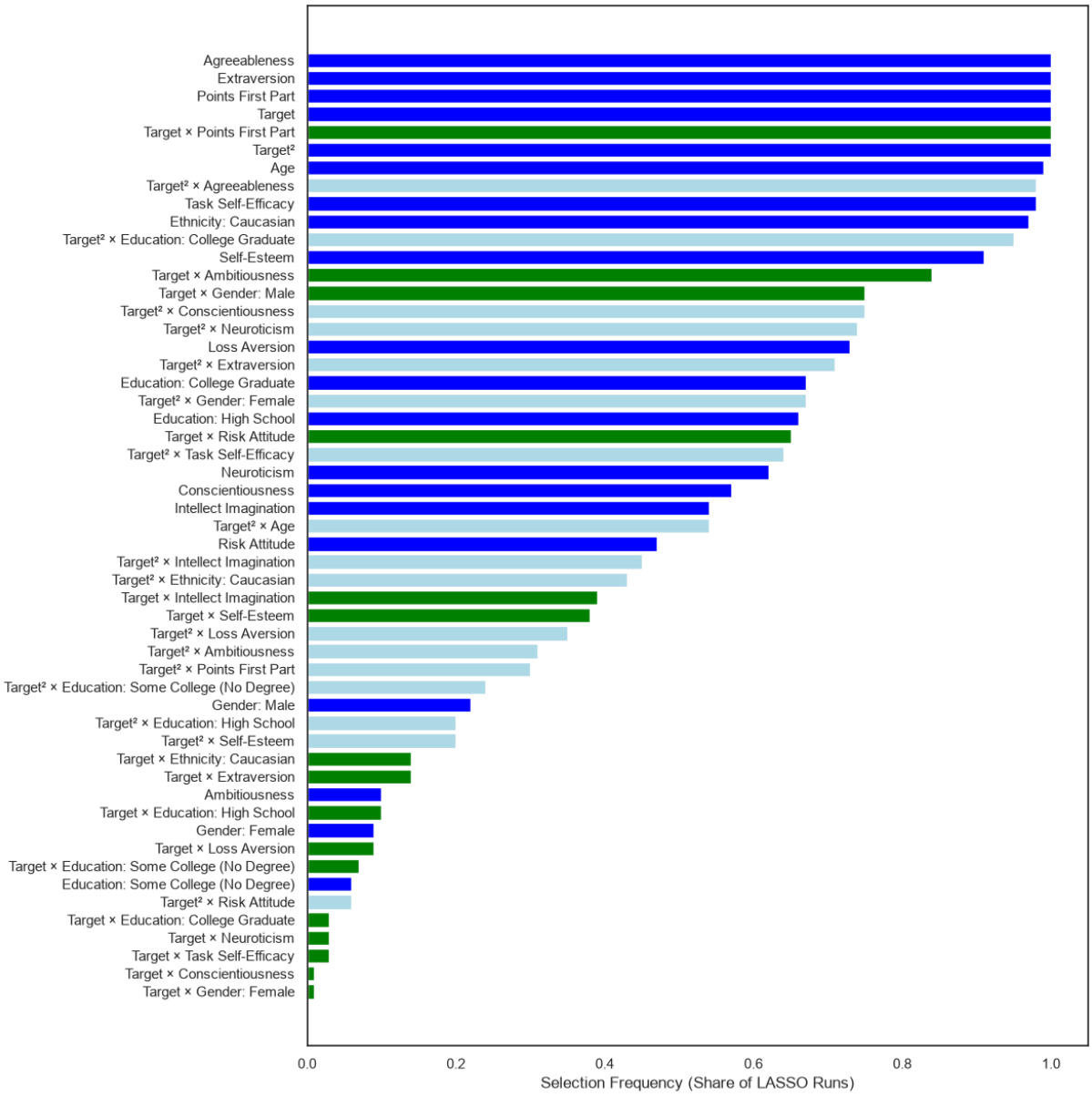

Figure 3: Frequency of the Variables Being Chosen in Lasso

Note: This figure shows how frequently the variables were chosen across 100 seeds that we used for the model computation (1.0 on x-axis means that variables were chosen in every model). Interaction terms with Target are marked with green, and interaction terms with Target^2 are marked with light blue. The features that were never selected by Lasso are Target × Agreeableness, Target × Age, Target^2× Gender: Male.

To investigate in more detail how the optimal target level varies with personal characteristics, we construct partial dependence plots that show the average predicted performance at each target level for different values of a given characteristic. To do so, we use two complementary approaches, each addressing a distinct question:

The first approach compares predicted target–response curves across observed subsamples. For example, we compare the curves for male and female participants to assess whether their predicted responses to targets differ. Because these comparisons are based on the observed subsamples, they preserve correlations between the characteristic of interest and all other covariates. Any differences between the curves may therefore reflect not only gender itself but also other characteristics that differ systematically between male and female participants.

The second approach constructs counterfactual partial dependence plots. Here, we set the characteristic of interest to a specific value for every individual in the sample while leaving all other covariates unchanged. For example, we predict target–response curves once after coding all participants as male and once after coding all participants as female. Comparing these curves isolates the difference in predicted performance attributable to changing the focal characteristic within the estimated model, rather than differences in the overall composition of the observed subsamples.

We implement both approaches and obtain very similar results. Figure 4 presents the counterfactual partial dependence plots, while Figure 6 in the Appendix reports the corresponding results based on the subsample approach.[14] For several characteristics, including risk attitude, loss aversion, self-esteem, intellect/imagination, age, task self-efficacy, and ethnicity, we find no meaningful heterogeneity in the predicted optimal target. For other characteristics, such as education, agreeableness, conscientiousness, extraversion, neuroticism or prior performance, the

[14]The corresponding plots based on the Random Forest model are reported in Figures 8 and 9 in the Appendix. Note that we use the non-parametric optimal common target of 2,700 as a benchmark here. This benchmark is more closely aligned with the flexibility of the Random Forest model because it does not rely on the quadratic functional-form restrictions imposed in the OLS analysis.

estimated optimum varies to some extent across individuals. While they vary in the direction also found in prior literature as discussed in Section 2[15], even in these cases, the predicted performance curves are relatively flat around the optimum, indicating that moderate deviations from the individually optimal target have only limited consequences for predicted performance.

Table 3: Performance loss from using a target of 2800 instead of the PDP-optimal target (Post-LASSO with Target and Target^2 interactions).

Variable Name	Avg Loss (Points)	Min Loss (Points)	Max Loss (Points)	Avg Loss (%)	Min Loss (%)	Max Loss (%)
Agreeableness	2.460	0.070	5.570	0.121	0.004	0.276
Points First Part	1.570	0.000	3.320	0.083	0.000	0.227
Education	1.040	0.000	2.190	0.050	0.000	0.105
Extraversion	0.960	0.000	2.160	0.047	0.000	0.108
Conscientiousness	0.340	0.000	0.910	0.017	0.000	0.045
Neuroticism	0.290	0.000	0.940	0.014	0.000	0.046
Ambitiousness	0.160	0.000	0.770	0.008	0.000	0.037
Gender	0.130	0.000	0.390	0.006	0.000	0.019
Self-Esteem	0.090	0.000	0.440	0.004	0.000	0.021
Age	0.000	0.000	0.000	0.000	0.000	0.000
Ethnicity	0.000	0.000	0.000	0.000	0.000	0.000
Intellect Imagination	0.000	0.000	0.000	0.000	0.000	0.000
Loss Aversion	0.000	0.000	0.000	0.000	0.000	0.000
Risk Attitude	0.000	0.000	0.000	0.000	0.000	0.000
Task Self-Efficacy	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table reports the average performance loss from using the optimal target (i.e., the target that maximizes predicted performance in the Post-Lasso model) relative to using the best average target (2800). The first three columns show the loss in absolute performance points, while the last three columns report relative losses. Each entry corresponds to a scenario in which a specific variable or personality trait is held fixed at a given value, and model predictions are computed across all possible goal levels, as in Figure 4.

To further assess the economic relevance of these differences for the different traits, we again compute the performance loss associated with assigning the average best target to everyone rather than assigning optimal individualized targets based on the respective individual characteristic. Specifically, we define the performance loss as the difference between predicted performance at the individual optimal target levels under the characteristic levels fixed in Figure 4 and the predicted performance at the benchmark target of 2,800, which corresponds to the best common target in the sample. Summary statistics for these performance losses are reported in Table 3. As the table shows, the average predicted performance loss from assigning the common optimal

[15]Higher levels of extraversion, conscientiousness and prior performance are associated with a larger predicted best target, while higher levels of agreeableness or neuroticism lead to a lower predicted best target (Locke and Latham [2002], Judge and Ilies [2002], Judge and Zapata [2015]) (see Figure 4) .

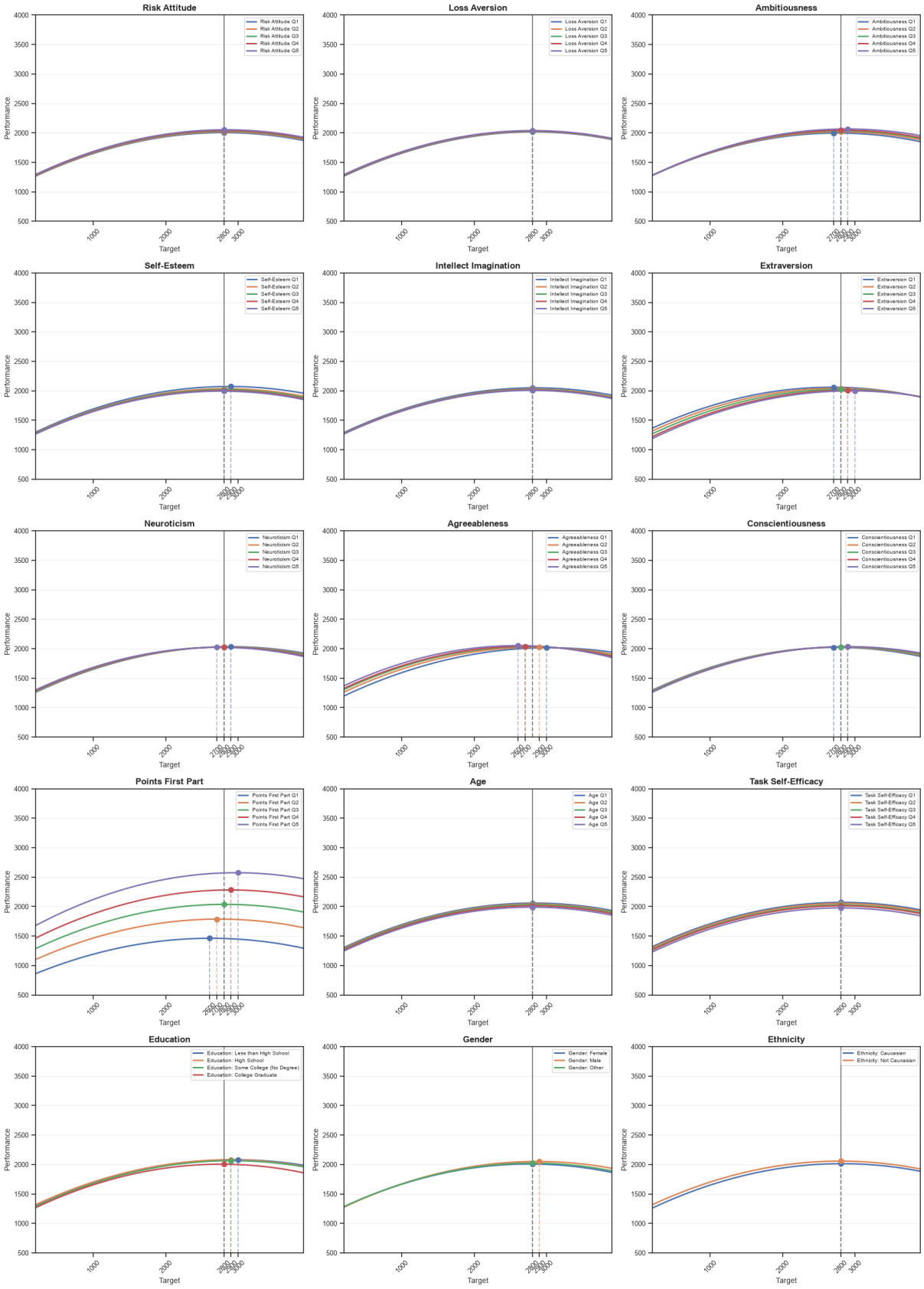

Figure 4: Partial Dependence Plot for Post-Lasso Model

Note: This figure presents partial dependence plots for predictions from the Post-Lasso model. The y-axis shows predicted performance, while the x-axis represents the target level. Each line corresponds to a scenario in which a specific variable or personality trait is held fixed at a given value, and model predictions are computed across all target levels. The points marked with dots indicate the peak of each curve, i.e., the target level at which predicted performance is maximized.

target of 2,800 points rather than the optimal target implied by each counterfactual characteristic value ranges from 0.09 points, or 0.004%, to 2.46 points, or 0.121%. Even the largest predicted performance loss, observed for the first quintile of agreeableness, amounts to only 5.57 points, or 0.28%.[16]

5 Conclusion

This paper examines whether personalizing performance target levels tailored to individual workers can meaningfully increase performance in a large scale online experiment. First, we replicate the pronounced inverted U-shaped relationship between target difficulty and performance: targets that are too low fail to stimulate effort, whereas targets that are too ambitious reduce performance. Thus, target difficulty clearly matters. However, our central result is that while average performance levels vary widely across individuals, the performance-maximizing targets vary little. Although some characteristics interact with target difficulty, the predicted optima cluster closely around the best common target, and the estimated gains from personalization are economically very small. Heterogeneity exists, but it appears not sufficiently actionable to justify individualized target assignment in this setting.

This result has important practical implications. From a design perspective, the limited heterogeneity in optimal targets is good news. It suggests that organizations can obtain most of the benefits of target setting by choosing a well-calibrated common target rather than by developing complex individualized target-setting systems. Such systems require data collection, model validation, monitoring, and explanation. They may also raise fairness concerns if comparable

[16]Performance loss for the subsample approach of building target-response curves for the partial dependence plots is presented in Table 4 in the Appendix. For this approach, the average predicted performance loss also ranges from 0.05 points to 2.29 points. The maximum predicted performance loss is 7.74 points or 0.375% of predicted performance under the optimal target. Same tables for performance losses for the Random Forest model can also be found in Appendix Tables 5 and 6. They show even smaller predicted losses when assigning the common best target instead of the individual one. Note that for the Random Forest model, we use the non-parametric optimal common target of 2,700 as a benchmark, since the greater functional flexibility of the non-parametric approach is more closely aligned with that of the Random Forest model than the parametric alternative.

workers receive different targets, as well as governance concerns about which characteristics are used and how algorithmic recommendations are controlled. When performance gains from personalization are small, as shown in our setting, a uniform target is not only simpler to implement but may also be easier to justify.

These findings should not be interpreted as evidence against personalization more generally. Instead, they show that its value depends on whether observable heterogeneity meaningfully changes the optimal managerial action. This distinction is particularly clear when comparing our findings to Opitz et al. [2025]. In the same task and a similar participant pool, tailoring the type of incentive generated sizable gains, whereas tailoring the level of a target generates negligible gains. Choosing among qualitatively different instruments may therefore be a first-order managerial decision, whereas fine-tuning the intensity of an already chosen instrument may be second order.

Several limitations remain. First, the study uses a specific real-effort task, which allows for clean random assignment and objective outcome measurement but abstracts from many features of organizational target setting, such as long-term employment relationships, social comparison, and managerial discretion. Although the task is well established and responsive to incentives (Opitz et al. [2025]), the extent to which our findings generalize to more complex workplace settings remains uncertain. Second, we study a single incentive structure in which reaching the target triggers a fixed bonus. Other arrangements, such as piecewise-linear rewards, penalties, team targets, repeated target revisions, or longer-term target cycles, may generate different patterns of heterogeneity and different returns to personalization. Finally, our assignment rules personalize targets only on the basis of the characteristics observed in the study. Additional information available in organizational settings could potentially identify further heterogeneity in optimal target difficulty.

Overall, this study shows that the value of individualizing management controls depends on the specific control decision. Although personalization can improve outcomes in some settings, it need not generate meaningful gains in others where a "one-size-fits-all" solution may work surprisingly well and come with lower transaction costs and social distortions. Future research should therefore identify the conditions under which the benefits of individualized controls are sufficiently large to outweigh the associated implementation, governance, and fairness costs.

References

Amir, O., and D. Ariely. Resting on laurels: the effects of discrete progress markers as subgoals on task performance and preferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34 (2008): 1158–1171.

Andreoni, J.; M. Callen; K. Hussain; M. Y. Khan; and C. Sprenger. Using preference estimates to customize incentives: an application to polio vaccination drives in Pakistan. *Journal of the European Economic Association* 21 (2023): 1428–1477.

Arnold, M. C., and M. Artz. Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. *Accounting, Organizations and Society* 40 (2015): 61–77.

Asad, S. A.; R. Banerjee; and J. Bhattacharya. Do workers discriminate against their out-group employers? Evidence from an online platform economy. *Journal of Economic Behavior & Organization* 216 (2023): 221–242.

Bandura, A. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 84 (1977): 191–215.

Bandura, A. The Role of Self-Efficacy in Goal-Based Motivation. In *New Developments in Goal Setting and Task Performance*, edited by E. A. Locke, and G. P. Latham, 147–157. New York, NY: Routledge. 2013.

Baron, R. A.; B. A. Mueller; and M. T. Wolfe. Self-efficacy and entrepreneurs' adoption of unattainable goals: The restraining effects of self-control. *Journal of Business Venturing* 31 (2016): 55–71.

Barrick, M. R., and M. K. Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology* 44 (1991): 1–26.

Becker-Peth, M.; E. Katok; and U. W. Thonemann. Designing buyback contracts for irrational but predictable newsvendors. *Management Science* 59 (2013): 1800–1816.

Belloni, A., and V. Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19 (2013): 521–547.

Berger, J., and D. Pope. Can losing lead to winning? *Management Science* 57 (2011): 817–827.

Danz, D.; L. Vesterlund; and A. J. Wilson. Belief elicitation and behavioral incentive compatibility. *American Economic Review* 112 (2022): 2851–2883.

De Quidt, J.; J. Haushofer; and C. Roth. Measuring and bounding experimenter demand. *American Economic Review* 108 (2018): 3266–3302.

Dekker, H. C.; T. Groot; and M. Schoute. Determining performance targets. *Behavioral Research in Accounting* 24 (2012): 21–46.

DellaVigna, S., and D. Pope. What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies* 85 (2018): 1029–1069.

Dinhof, K.; S. Neo; I. Bertram; R. Bouwman; N. De Boer; G. Szydlowski; J. Willems; and L. Tummers. The threat of appearing lazy, inefficient, and slow? Stereotype threat in the public sector. *Public Management Review* 26 (2024): 1941–1962.

Dohmen, T.; A. Falk; D. Huffman; U. Sunde; J. Schupp; and G. G. Wagner. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9 (2011): 522–550.

Donnellan, M. B.; F. L. Oswald; B. M. Baird; and R. E. Lucas. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment* 18 (2006): 192–203.

Dossett, D. L.; G. P. Latham; and T. R. Mitchell. Effects of assigned versus participatively set goals, knowledge of results, and individual differences on employee behavior when goal difficulty is held constant. *Journal of Applied Psychology* 64 (1979): 291–298.

Earley, P. C.; T. Connolly; and G. Ekegren. Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology* 74 (1989): 24–33.

Erez, A., and T. A. Judge. Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology* 86 (2001): 1270–1279.

Erez, M. Feedback: A necessary condition for the goal setting-performance relationship. *Journal of Applied Psychology* 62 (1977): 624–627.

Erez, M., and I. Zidon. Effect of goal acceptance on the relationship of goal difficulty to performance. *Journal of Applied Psychology* 69 (1984): 69–78.

Gopalakrishnan, M.; T. Libby; J. A. Samuels; and D. Swenson. The effect of cost goal specificity and new product development process on cost reduction performance. *Accounting, Organizations and Society* 42 (2015): 1–11.

Gordon, M. L.; T. Althoff; and J. Leskovec. Goal-setting and Achievement in Activity Tracking Apps: A Case Study of MyFitnessPal. In *Proceedings of the 2019 World Wide Web Conference*, 571–582. New York, NY, USA: Association for Computing Machinery. 2019.

Gosnell, G. K.; J. A. List; and R. D. Metcalfe. The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy* 128 (2020): 1195–1233.

Heath, C.; R. P. Larrick; and G. Wu. Goals as reference points. *Cognitive Psychology* 38 (1999): 79–109.

Hirschi, A., and D. Spurk. Striving for success: Towards a refined understanding and measurement of ambition. *Journal of Vocational Behavior* 127 (2021): 103577.

Hitsch, G. J.; S. Misra; and W. W. Zhang. Heterogeneous treatment effects and optimal targeting policy evaluation. *Quantitative Marketing and Economics* 22 (2024): 115–168.

Judge, T. A., and R. Ilies. Relationship of personality to performance motivation: a meta-analytic review. *Journal of Applied Psychology* 87 (2002): 797–807.

Judge, T. A., and C. P. Zapata. The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal* 58 (2015): 1149–1179.

Kim, S.; M. Matějka; and J. Park. Economic determinants and consequences of performance target difficulty. *The Accounting Review* 98 (2023): 361–387.

Knight, D.; C. C. Durham; and E. A. Locke. The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. *Academy of Management Journal* 44 (2001): 326–338.

Larrick, R. P.; C. Heath; and G. Wu. Goal-induced risk taking in negotiation and decision making. *Social Cognition* 27 (2009): 342–364.

Latham, G. P., and L. M. Saari. Importance of supportive relationships in goal setting. *Journal of Applied Psychology* 64 (1979): 151–156.

Latham, G. P.; D. C. Winters; and E. A. Locke. Cognitive and motivational effects of participation: A mediator study. *Journal of Organizational Behavior* 15 (1994): 49–63.

Levy, P. E., and A. H. Baumgardner. Effects of self-esteem and gender on goal choice. *Journal of Organizational Behavior* 12 (1991): 529–541.

Li, J.; L. Chai; O. Nordstrom; C. Tangpong; and K.-T. Hung. Development of a Loss Aversion Scale. *Journal of Managerial Issues* 33 (2021): 69–89.

Lipman, S. A.; N. W. Boderie; J. V. Been; and H. Van Kippersluis. Deposit? Yes, please! The effect of different modes of assigning reward-and deposit-based financial incentives on effort. *Behavioural Public Policy* 10 (2026): 13–41.

List, J. A.; A. M. Shaikh; and A. Vayalinkal. Multiple testing with covariate adjustment in experimental economics. *Journal of Applied Econometrics* 38 (2023): 920–939.

List, J. A.; A. M. Shaikh; and Y. Xu. Multiple hypothesis testing in experimental economics. *Experimental Economics* 22 (2019): 773–793.

Locke, E. A., and G. P. Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57 (2002): 705–717.

Locke, E. A.; K. N. Shaw; L. M. Saari; and G. P. Latham. Goal setting and task performance: 1969–1980. *Psychological Bulletin* 90 (1981): 125–152.

Locke, E. A., and G. P. Latham. A Theory of Goal Setting & Task Performance. Englewood Cliffs, NJ: Prentice-Hall. 1990.

Locke, E. A., and G. P. Latham. New Developments in Goal Setting and Task Performance. New York, NY: Routledge. 2013.

Locke, E. A., and G. P. Latham. Breaking the Rules: A Historical Overview of Goal-Setting Theory. In *Advances in Motivation Science*, vol. 2, edited by A. J. Elliot, 99–126. Elsevier. 2015.

Marginson, D., and S. Ogden. Coping with ambiguity through the budget: the positive effects of budgetary targets on managers' budgeting behaviours. *Accounting, Organizations and Society* 30 (2005): 435–456.

Matějka, M. Target setting in multi-divisional organizations. *Journal of Management Accounting Research* 30 (2018): 13–27.

Matějka, M.; M. D. Mahlendorf; and U. Schäffer. The ratchet effect: Theory and empirical evidence. *Management Science* 70 (2024): 128–142.

Matějka, M., and K. Ray. Balancing difficulty of performance targets: Theory and evidence. *Review of Accounting Studies* 22 (2017): 1666–1697.

Miner, J. B. The validity and usefulness of theories in an emerging organizational science. *Academy of Management Review* 9 (1984): 296–306.

Murakami, K.; H. Shimada; Y. Ushifusa; and T. Ida. Heterogeneous treatment effects of nudge and rebate: Causal machine learning in a field experiment on electricity conservation. *International Economic Review* 63 (2022): 1779–1803.

Opitz, S.; D. Sliwka; T. Vogelsang; and T. Zimmermann. The Algorithmic Assignment of Incentive Schemes. *Management Science* 71 (2025): 1546–1563.

Pilegge, A. J., and R. Holtz. The effects of social identity on the self-set goals and task performance of high and low self-esteem individuals. *Organizational Behavior and Human Decision Processes* 70 (1997): 17–26.

Pinder, C. C. Work motivation: Theory, issues, and applications. Glenview, IL: Scott, Foresman and Company. 1984.

Rosenberg, M. Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press. 2015.

Schrauger, J. S., and S. E. Rosenberg. Self-esteem and the effects of success and failure feedback on performance. *Journal of Personality* 38 (1970): 404–417.

Shchetkina, A., and R. Berman. When is heterogeneity actionable for personalization? *arXiv preprint arXiv:2411.16552* (2024).

Stajkovic, A. D., and F. Luthans. Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin* 124 (1998): 240–261.

Winters, D., and G. P. Latham. The effect of learning versus outcome goals on a simple versus a complex task. *Group & Organization Management* 21 (1996): 236–250.

Yukl, G. A., and G. P. Latham. Interrelationships among employee participation, individual differences, goal difficulty, goal acceptance, goal instrumentality, and performance. *Personnel Psychology* 31 (1978): 305–323.

A Appendix: General

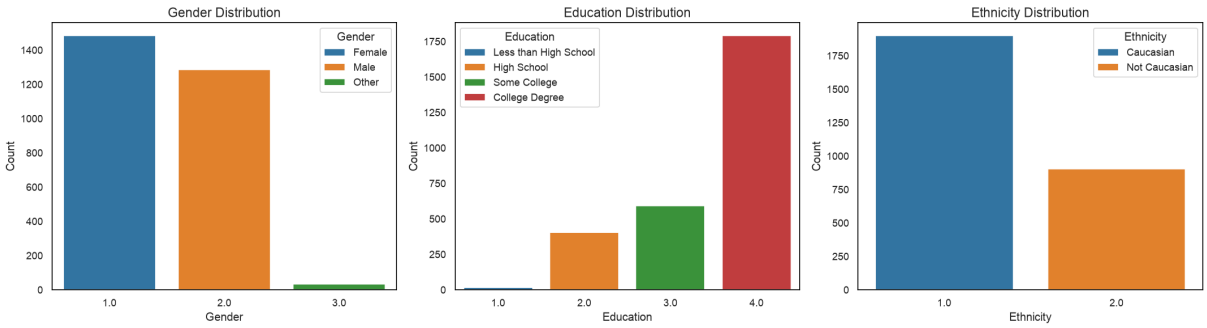

Figure 5: Demographics of the Participants
Note: This figure shows the demographic distribution of the participants.

Table 4: Performance loss from using a target of 2800 instead of the PDP-optimal target (Post-LASSO with Target and Target^2 interactions, Data Subsets Method).

Variable Name	Avg Loss (Points)	Min Loss (Points)	Max Loss (Points)	Avg Loss (%)	Min Loss (%)	Max Loss (%)
Agreeableness	2.290	0.000	7.740	0.112	0.000	0.375
Points First Part	2.170	0.000	6.160	0.107	0.000	0.236
Gender	1.150	0.320	2.020	0.056	0.015	0.096
Education	0.870	0.000	1.870	0.040	0.000	0.083
Task Self-Efficacy	0.560	0.000	2.300	0.024	0.000	0.097
Neuroticism	0.490	0.000	1.480	0.024	0.000	0.072
Ambitiousness	0.450	0.000	1.680	0.021	0.000	0.079
Risk Attitude	0.340	0.000	1.270	0.016	0.000	0.061
Conscientiousness	0.330	0.000	0.860	0.016	0.000	0.042
Extraversion	0.240	0.000	0.920	0.012	0.000	0.047
Intellect Imagination	0.210	0.000	0.580	0.010	0.000	0.029
Age	0.160	0.000	0.430	0.008	0.000	0.021
Loss Aversion	0.110	0.000	0.530	0.005	0.000	0.026
Self-Esteem	0.050	0.000	0.240	0.002	0.000	0.012
Ethnicity	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table reports the average performance loss from using the optimal target (i.e., the target that maximizes predicted performance in the Post-Lasso model) relative to using the best average target (2800). The first three columns show the loss in absolute performance points, while the last three columns report relative losses. Each entry corresponds to a scenario in which the data is filtered for the specific category of a variable or a particular quintile of a variable if it is continuous, and model predictions are computed across all goal levels, as in Figure 6.

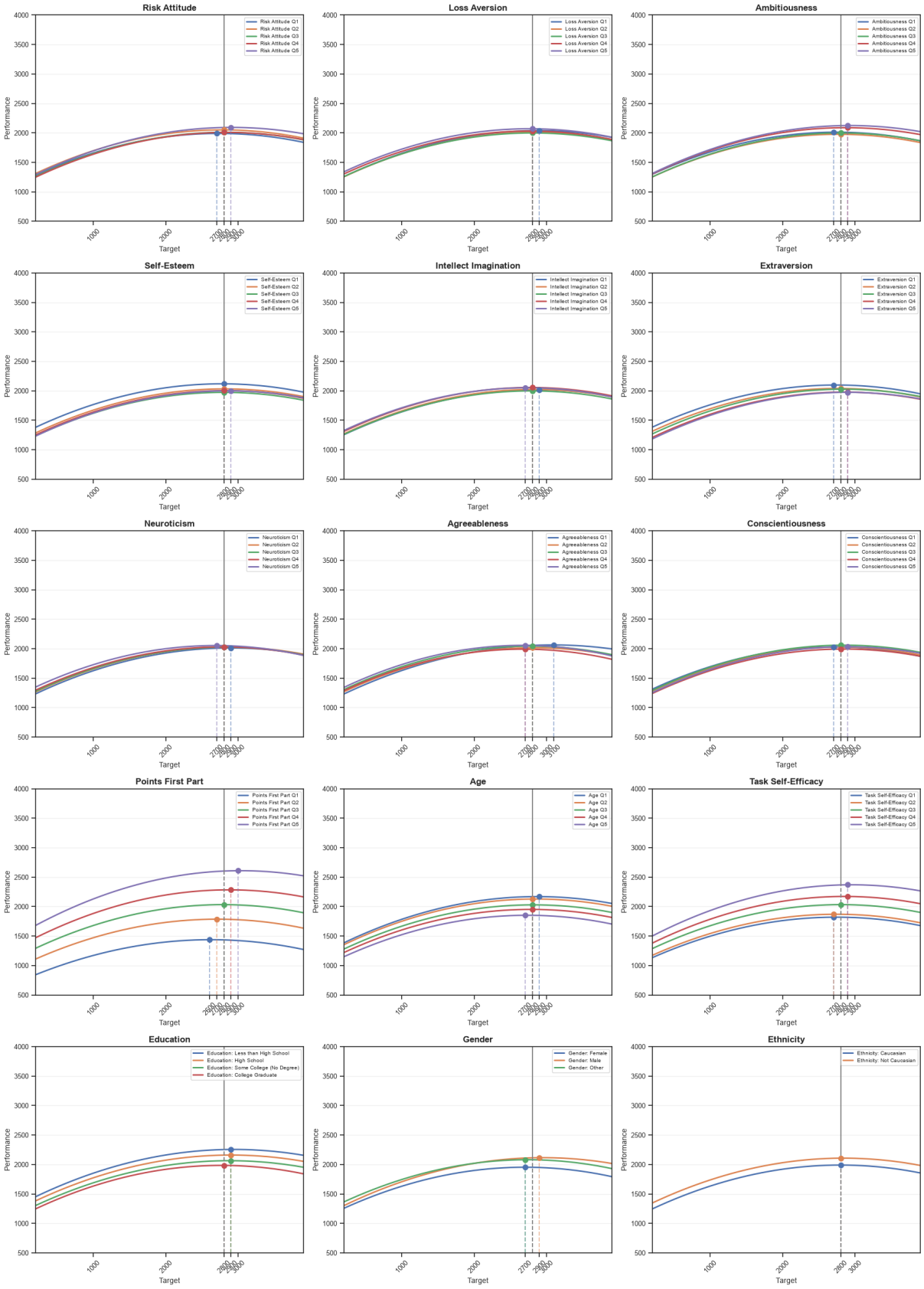

Figure 6: Partial Dependence Plot for Post-Lasso Model, Subset Method

Note: This figure presents partial dependence plots for predictions from the Post-Lasso model. The y-axis shows predicted performance, while the x-axis represents the target level. Each line corresponds to a scenario in which the data is filtered for the specific category of a variable or a particular quintile of a variable if it is continuous, and model predictions are computed across all target levels. The points marked with dots indicate the peak of each curve, i.e., the target level at which predicted performance is maximized.

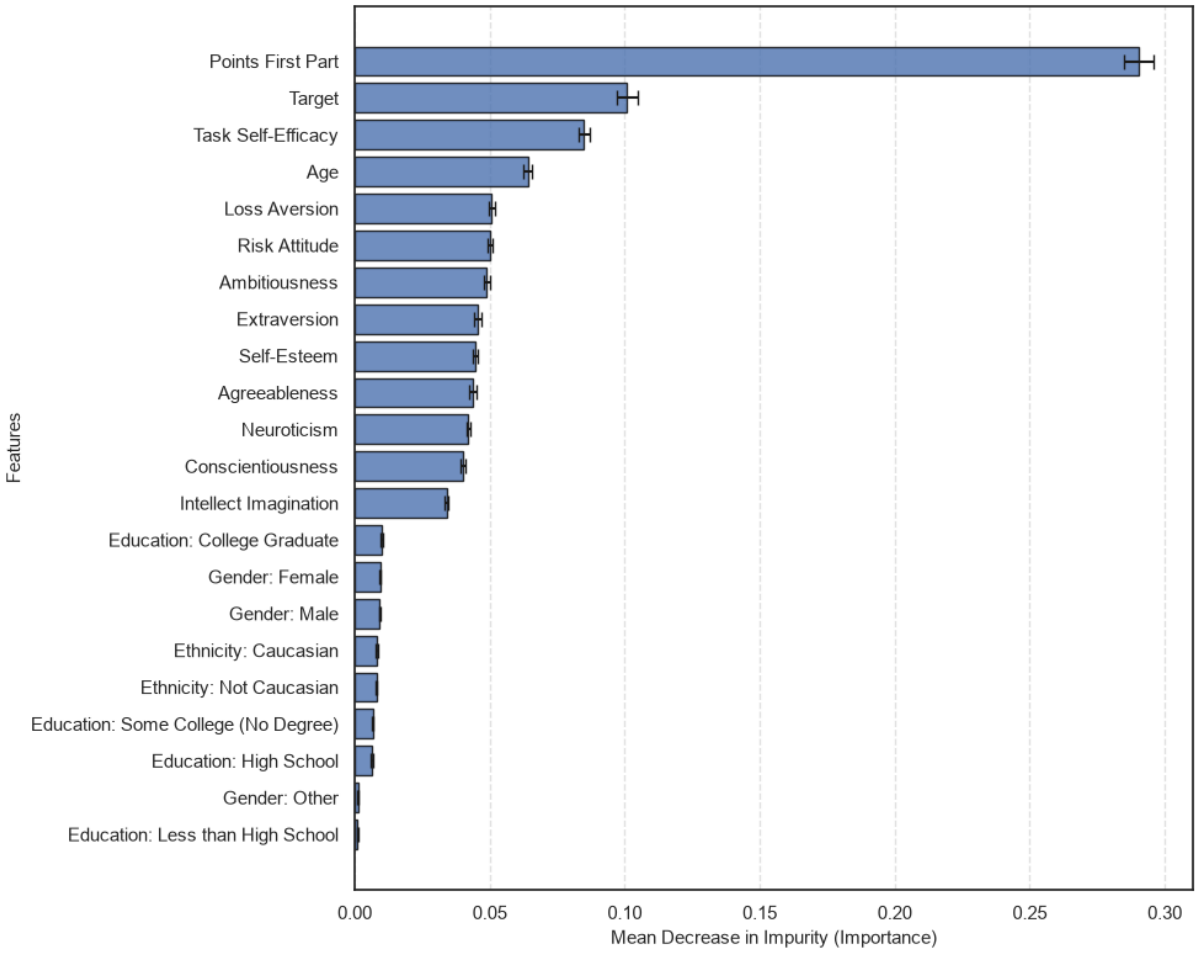

Figure 7: Variable Importance Plot for Random Forest Model
Note: This figure shows the variable importance for the Random Forest model.

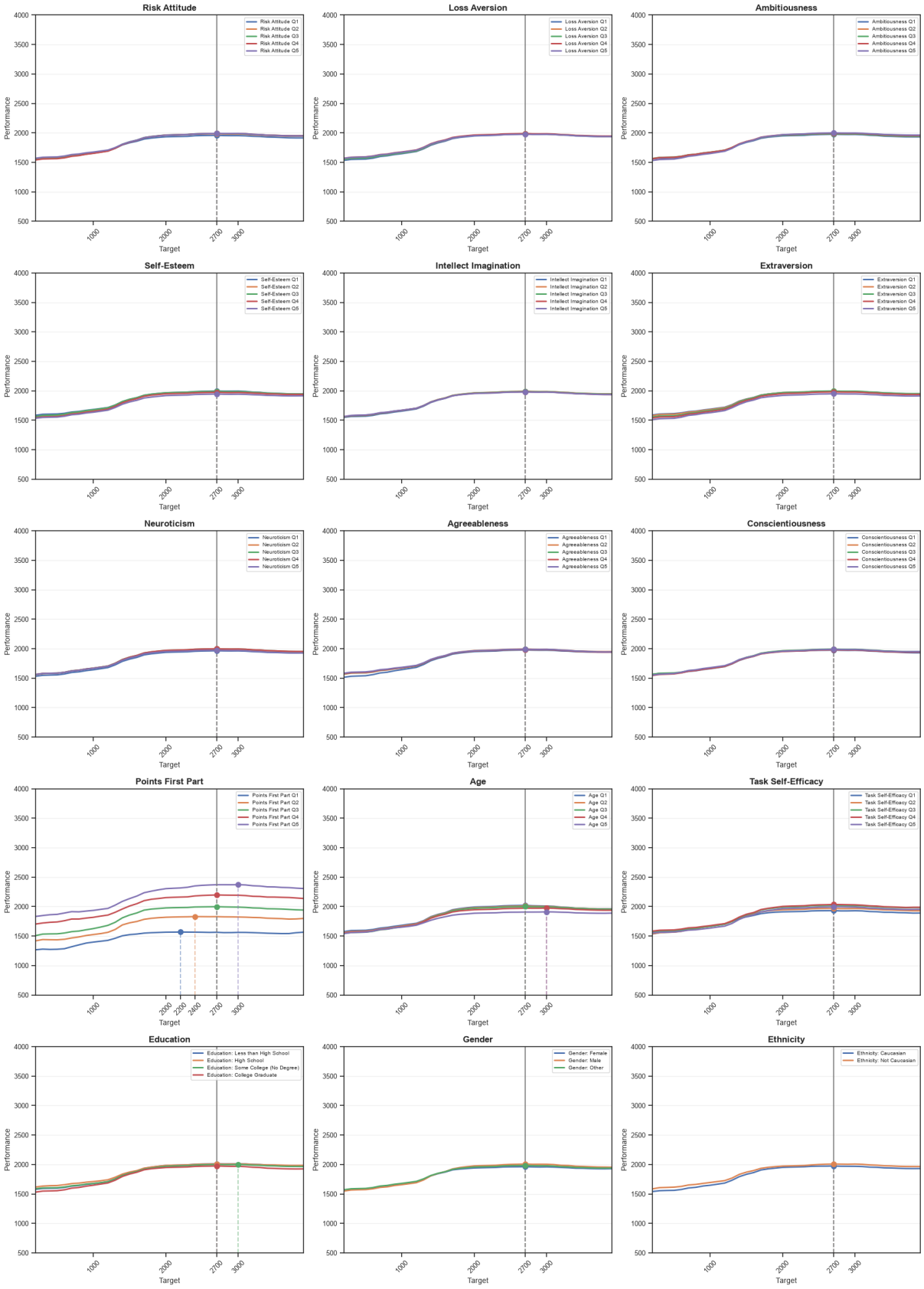

Figure 8: Partial Dependence Plot for Random Forest Model

Note: This figure presents partial dependence plots for predictions from the Random Forest model. The y-axis shows predicted performance, while the x-axis represents the target level. Each line corresponds to a scenario in which a specific variable or personality trait is held fixed at a given value, and model predictions are computed across all target levels. The points marked with dots indicate the peak of each curve, i.e., the target level at which predicted performance is maximized. We use the non-parametric optimal common target of 2,700 as a benchmark here. This benchmark is more closely aligned with the flexibility of the Random Forest model because it does not rely on the quadratic functional-form restrictions imposed in the OLS analysis.

Table 5: Performance loss from using a target of 2700 instead of the PDP-optimal target (Random Forest Model).

Variable Name	Avg Loss (Points)	Min Loss (Points)	Max Loss (Points)	Avg Loss (%)	Min Loss (%)	Max Loss (%)
Points First Part	1.210	0.000	4.140	0.071	0.000	0.264
Age	1.220	0.000	4.390	0.063	0.000	0.230
Education	0.050	0.000	0.210	0.003	0.000	0.010
Agreeableness	0.000	0.000	0.000	0.000	0.000	0.000
Ambitiousness	0.000	0.000	0.000	0.000	0.000	0.000
Conscientiousness	0.000	0.000	0.000	0.000	0.000	0.000
Ethnicity	0.000	0.000	0.000	0.000	0.000	0.000
Extraversion	0.000	0.000	0.000	0.000	0.000	0.000
Gender	0.000	0.000	0.000	0.000	0.000	0.000
Intellect Imagination	0.000	0.000	0.000	0.000	0.000	0.000
Loss Aversion	0.000	0.000	0.000	0.000	0.000	0.000
Neuroticism	0.000	0.000	0.000	0.000	0.000	0.000
Risk Attitude	0.000	0.000	0.000	0.000	0.000	0.000
Self-Esteem	0.000	0.000	0.000	0.000	0.000	0.000
Task Self-Efficacy	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table reports the average performance loss from using the optimal target (i.e., the target that maximizes predicted performance in the Random Forest model) relative to using the nonparametrically estimated optimal common target (2700). We use the non-parametric optimal common target of 2,700 as a benchmark here because it is more closely aligned with the flexibility of the random forest model since it does not rely on the quadratic functional-form restrictions imposed in the OLS analysis. The first three columns show the loss in absolute performance points, while the last three columns report relative losses. Each entry corresponds to a scenario in which a specific variable or personality trait is held fixed at a given value, and model predictions are computed across all possible goal levels, as in Figure 8.

Table 6: Performance loss from using a target of 2700 instead of the PDP-optimal target (Random Forest with Target and Target^2 interactions, Data Subsets Method).

Variable Name	Avg Loss (Points)	Min Loss (Points)	Max Loss (Points)	Avg Loss (%)	Min Loss (%)	Max Loss (%)
Points First Part	2.680	0.000	9.290	0.171	0.000	0.630
Age	0.850	0.000	3.030	0.047	0.000	0.170
Education	0.380	0.000	1.530	0.017	0.000	0.068
Task Self-Efficacy	0.150	0.000	0.530	0.007	0.000	0.023
Risk Attitude	0.000	0.000	0.010	0.000	0.000	0.000
Agreeableness	0.000	0.000	0.000	0.000	0.000	0.000
Ambitiousness	0.000	0.000	0.000	0.000	0.000	0.000
Conscientiousness	0.000	0.000	0.000	0.000	0.000	0.000
Ethnicity	0.000	0.000	0.000	0.000	0.000	0.000
Extraversion	0.000	0.000	0.000	0.000	0.000	0.000
Gender	0.000	0.000	0.000	0.000	0.000	0.000
Intellect Imagination	0.000	0.000	0.000	0.000	0.000	0.000
Loss Aversion	0.000	0.000	0.000	0.000	0.000	0.000
Neuroticism	0.000	0.000	0.000	0.000	0.000	0.000
Self-Esteem	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table reports the average performance loss from using the optimal target (i.e., the target that maximizes predicted performance in the Random Forest model) relative to using the nonparametrically estimated optimal common target (2700). We use the non-parametric optimal common target of 2,700 as a benchmark here because it is more closely aligned with the flexibility of the random forest model since it does not rely on the quadratic functional-form restrictions imposed in the OLS analysis. The first three columns show the loss in absolute performance points, while the last three columns report relative losses. Each entry corresponds to a scenario in which the data is filtered for the specific category of a variable or a particular quintile of a variable if it is continuous, and model predictions are computed across all goal levels, as in Figure 9.

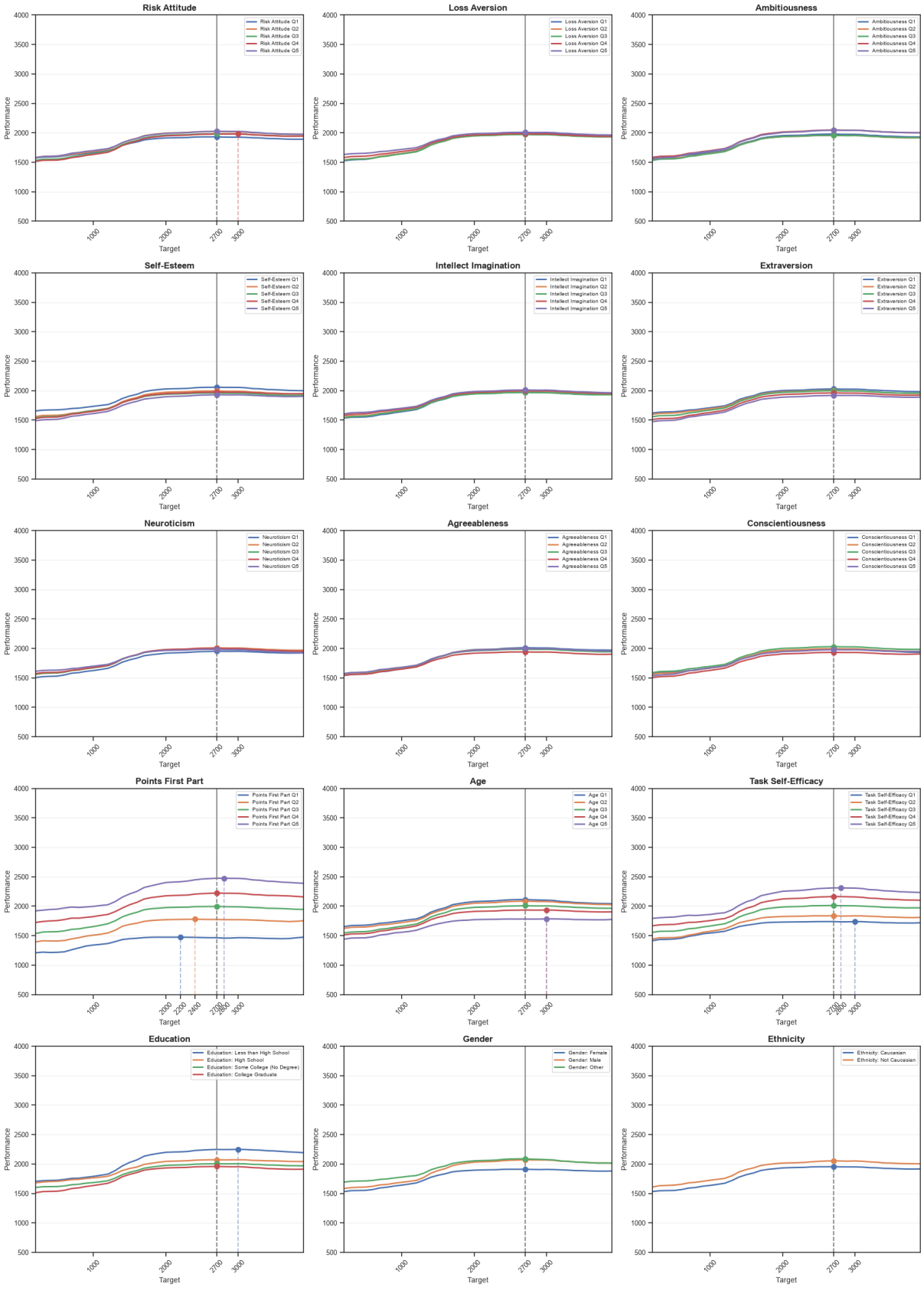

Figure 9: Partial Dependence Plot for Random Forest Model, Subset Method

Note: This figure presents partial dependence plots for predictions from the Random Forest model. The y-axis shows predicted performance, while the x-axis represents the target level. Each line corresponds to a scenario in which the data is filtered for the specific category of a variable or a particular quintile of a variable if it is continuous, and model predictions are computed across all target levels. The points marked with dots indicate the peak of each curve, i.e., the target level at which predicted performance is maximized. We use the non-parametric optimal common target of 2,700 as a benchmark here because it is more closely aligned with the flexibility of the Random Forest model since it does not rely on the quadratic functional-form restrictions imposed in the OLS analysis.

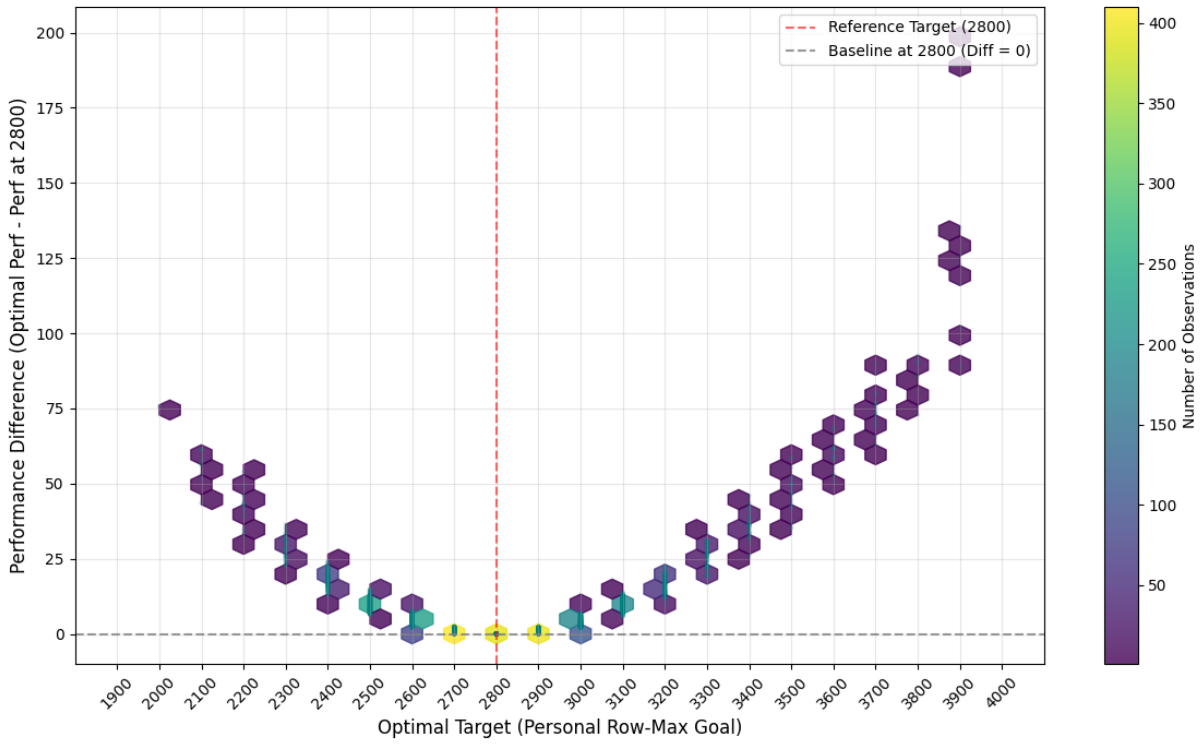

Figure 10: Predicted Performance Difference From Using a Target of 2,800 instead of the Optimal Target

Note: This figure shows the difference in predicted performance under optimal target relative to the one under the target level of 2,800 for all observations in the dataset. Each point represents the predicted performance difference under a given target. The predicted performance under the 2,800 target is set as the reference point and is therefore marked as zero. The average performance difference is 8.37 points, and the standard deviation is 14.31 points.

Table 7: Regression Results: Points Second Part (Models 1–8)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Prior Perf.	Prior Qs	Ethnicity	Education	Agree.	Extra.	Neuro.	Consc.
Constant	430.848***	781.259***	1261.238***	-256.489	1199.204***	1191.291***	1193.674***	1193.644***
	(139.462)	(108.672)	(107.602)	(238.506)	(62.998)	(62.780)	(63.501)	(63.434)
Target	0.436***	0.504***	0.586***	1.922***	0.588***	0.594***	0.594***	0.594***
	(0.144)	(0.120)	(0.104)	(0.289)	(0.062)	(0.061)	(0.062)	(0.062)
Target^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Points First Part	0.430***							
	(0.083)							
Target × Points First Part	0.000							
	(0.000)							
Target^2 × Points First Part	-0.000							
	(0.000)							
Prior Perf. Q2		233.677						
		(153.482)						
Prior Perf. Q3		399.015**						
		(157.209)						
Prior Perf. Q4		466.289***						
		(176.455)						
Prior Perf. Q5		822.119***						
		(177.968)						
Target × Prior Perf. Q2		0.041						
		(0.165)						
Target^2 × Prior Perf. Q2		0.000						
		(0.000)						
Target × Prior Perf. Q3		0.148						
		(0.163)						
Target^2 × Prior Perf. Q3		-0.000						
		(0.000)						
Target × Prior Perf. Q4		0.190						
		(0.174)						
Target^2 × Prior Perf. Q4		-0.000						
		(0.000)						
Target × Prior Perf. Q5		0.154						
		(0.177)						
Target^2 × Prior Perf. Q5		-0.000						
		(0.000)						
Ethnicity: Caucasian			-107.524					
			(133.385)					
Target × Ethnicity: Caucasian			0.022					
			(0.129)					
Target^2 × Ethnicity: Caucasian			-0.000					
			(0.000)					
Education: High School				1925.204***				
				(285.733)				
Education: Some College (No Degree)				1554.245***				
				(271.935)				
Education: College Graduate				1298.587***				
				(251.500)				
Target × Education: High School				-1.707***				
				(0.327)				
Target^2 × Education: High School				0.000***				
				(0.000)				
Target × Education: Some College (No Degree)				-1.420***				
				(0.316)				
Target^2 × Education: Some College (No Degree)				0.000***				
				(0.000)				
Target × Education: College Graduate				-1.206***				
				(0.300)				
Target^2 × Education: College Graduate				0.000***				
				(0.000)				
Agreeableness					103.592			
					(65.742)			
Target × Agreeableness					-0.064			
					(0.065)			
Target^2 × Agreeableness					0.000			
					(0.000)			
Extraversion						-158.054***		
						(60.850)		
Target × Extraversion						0.098		
						(0.060)		
Target^2 × Extraversion						-0.000		
						(0.000)		
Neuroticism							82.958	
							(64.803)	
Target × Neuroticism							-0.036	
							(0.063)	
Target^2 × Neuroticism							0.000	
							(0.000)	
Conscientiousness								-36.120
								(67.646)
Target × Conscientiousness								0.002
								(0.064)
Target^2 × Conscientiousness								0.000
								(0.000)
Observations	2804	2804	2804	2804	2804	2804	2804	2804
R-squared	0.358	0.321	0.059	0.066	0.055	0.059	0.055	0.054

This table reports OLS estimates of the relationship between performance and assigned targets. Each model includes a quadratic specification in the target level and its interactions with a single participant characteristic. Continuous characteristics enter linearly, while categorical variables are represented by indicator variables with omitted reference categories. For variables grouped into quintiles, the first quintile serves as the reference category. Robust standard errors are reported in parentheses. Statistical significance is denoted by $* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table 8: Regression Results: Points Second Part (Models 9–16)

Variable	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Amb.	Self-Est.	Risk	Loss	Intellect	Gender	Age	Self-Eff. Qs
Constant	1191.052***	1187.509***	1191.666***	1205.682***	1192.615***	1860.459***	1132.178***	1175.553***
	(63.307)	(62.892)	(63.441)	(63.416)	(63.759)	(445.563)	(203.238)	(123.150)
Target	0.596***	0.598***	0.596***	0.584***	0.594***	0.500	1.090***	0.463***
	(0.062)	(0.061)	(0.062)	(0.062)	(0.062)	(0.461)	(0.201)	(0.125)
Target^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ambitiousness	-98.616							
	(67.448)							
Target × Ambitiousness	0.094							
	(0.064)							
Target^2 × Ambitiousness	-0.000							
	(0.000)							
Self-Esteem		-110.910*						
		(62.088)						
Target × Self-Esteem		0.041						
		(0.061)						
Target^2 × Self-Esteem		-0.000						
		(0.000)						
Risk Attitude			-114.567					
			(69.808)					
Target × Risk Attitude			0.108					
			(0.066)					
Target^2 × Risk Attitude			-0.000					
			(0.000)					
Loss Aversion				149.873**				
				(65.812)				
Target × Loss Aversion				-0.118*				
				(0.064)				
Target^2 × Loss Aversion				0.000				
				(0.000)				
Intellect Imagination					45.900			
					(69.245)			
Target × Intellect Imagination					-0.026			
					(0.066)			
Target^2 × Intellect Imagination					0.000			
					(0.000)			
Gender: Female						-589.159		
						(453.131)		
Gender: Male						-755.579*		
						(456.149)		
Target × Gender: Female						-0.044		
						(0.468)		
Target^2 × Gender: Female						0.000		
						(0.000)		
Target × Gender: Male						0.246		
						(0.471)		
Target^2 × Gender: Male						0.000		
						(0.000)		
Age							1.526	
							(4.285)	
Target × Age							-0.011***	
							(0.004)	
Target^2 × Age							0.000***	
							(0.000)	
Task Self-Eff. Q2								8.614
								(173.446)
Task Self-Eff. Q3								18.801
								(169.259)
Task Self-Eff. Q4								218.515
								(228.562)
Task Self-Eff. Q5								-2.854
								(228.115)
Target × Task Self-Eff. Q2								-0.007
								(0.176)
Target^2 × Task Self-Eff. Q2								0.000
								(0.000)
Target × Task Self-Eff. Q3								0.147
								(0.167)
Target^2 × Task Self-Eff. Q3								-0.000
								(0.000)
Target × Task Self-Eff. Q4								0.142
								(0.221)
Target^2 × Task Self-Eff. Q4								-0.000
								(0.000)
Target × Task Self-Eff. Q5								0.349
								(0.217)
Target^2 × Task Self-Eff. Q5								-0.000
								(0.000)
Observations	2804	2804	2804	2804	2804	2804	2804	2804
R-squared	0.057	0.058	0.056	0.055	0.054	0.066	0.079	0.118

This table reports OLS estimates of the relationship between performance and assigned targets. Each model includes a quadratic specification in the target level and its interactions with a single participant characteristic. Continuous characteristics enter linearly, while categorical variables are represented by indicator variables with omitted reference categories. For variables grouped into quintiles, the first quintile serves as the reference category. Robust standard errors are reported in parentheses. Statistical significance is denoted by $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table 9: Regression Results: Points Second Part (Models 1–8)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Prior Perf.	Prior Qs	Ethnicity	Education	Agree.	Extra.	Neuro.	Consc.
Constant	430.848	781.259*	1261.238***	-256.489	1199.204***	1191.291***	1193.674***	1193.644***
	(139.462)	(108.672)	(107.602)	(238.506)	(62.998)	(62.780)	(63.501)	(63.434)
Target	0.436	0.504	0.586	1.922*	0.588**	0.594**	0.594**	0.594**
	(0.144)	(0.120)	(0.104)	(0.289)	(0.062)	(0.061)	(0.062)	(0.062)
Target^2	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Points First Part	0.430							
	(0.083)							
Target × Points First Part	0.000							
	(0.000)							
Target^2 × Points First Part	-0.000							
	(0.000)							
Prior Perf. Q2		233.677						
		(153.482)						
Prior Perf. Q3		399.015						
		(157.209)						
Prior Perf. Q4		466.289						
		(176.455)						
Prior Perf. Q5		822.119						
		(177.968)						
Target × Prior Perf. Q2		0.041						
		(0.165)						
Target^2 × Prior Perf. Q2		0.000						
		(0.000)						
Target × Prior Perf. Q3		0.148						
		(0.163)						
Target^2 × Prior Perf. Q3		-0.000						
		(0.000)						
Target × Prior Perf. Q4		0.190						
		(0.174)						
Target^2 × Prior Perf. Q4		-0.000						
		(0.000)						
Target × Prior Perf. Q5		0.154						
		(0.177)						
Target^2 × Prior Perf. Q5		-0.000						
		(0.000)						
Ethnicity: Caucasian			-107.524					
			(133.385)					
Target × Ethnicity: Caucasian			0.022					
			(0.129)					
Target^2 × Ethnicity: Caucasian			-0.000					
			(0.000)					
Education: High School				1925.204*				
				(285.733)				
Education: Some College (No Degree)				1554.245				
				(271.935)				
Education: College Graduate				1298.587				
				(251.500)				
Target × Education: High School				-1.707				
				(0.327)				
Target^2 × Education: High School				0.000				
				(0.000)				
Target × Education: Some College (No Degree)				-1.420				
				(0.316)				
Target^2 × Education: Some College (No Degree)				0.000				
				(0.000)				
Target × Education: College Graduate				-1.206				
				(0.300)				
Target^2 × Education: College Graduate				0.000				
				(0.000)				
Agreeableness					103.592			
					(65.742)			
Target × Agreeableness					-0.064			
					(0.065)			
Target^2 × Agreeableness					0.000			
					(0.000)			
Extraversion						-158.054		
						(60.850)		
Target × Extraversion						0.098		
						(0.060)		
Target^2 × Extraversion						-0.000		
						(0.000)		
Neuroticism							82.958	
							(64.803)	
Target × Neuroticism							-0.036	
							(0.063)	
Target^2 × Neuroticism							0.000	
							(0.000)	
Conscientiousness								-36.120
								(67.646)
Target × Conscientiousness								0.002
								(0.064)
Target^2 × Conscientiousness								0.000
								(0.000)
Observations	2804	2804	2804	2804	2804	2804	2804	2804
R-squared	0.358	0.321	0.059	0.066	0.055	0.059	0.055	0.054

This table reports the same OLS specifications as Tables 7 and 8. Reported p-values are adjusted for multiple hypothesis testing using the bootstrap step-down procedure of List et al. [2019], [2023], controlling the family-wise error rate across all estimated models. Robust standard errors are reported in parentheses. Statistical significance is denoted by $* \; p < 0.1$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table 10: Regression Results: Points Second Part (Models 9–16)

Variable	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Amb.	Self-Est.	Risk	Loss	Intellect	Gender	Age	Self-Eff. Qs
Constant	1191.052***	1187.509***	1191.666***	1205.682***	1192.615***	1860.459	1132.178	1175.553**
	(63.307)	(62.892)	(63.441)	(63.416)	(63.759)	(445.563)	(203.238)	(123.150)
Target	0.596**	0.598**	0.596**	0.584**	0.594**	0.500	1.090	0.463
	(0.062)	(0.061)	(0.062)	(0.062)	(0.062)	(0.461)	(0.201)	(0.125)
Target^2	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ambitiousness	-98.616							
	(67.448)							
Target × Ambitiousness	0.094							
	(0.064)							
Target^2 × Ambitiousness	-0.000							
	(0.000)							
Self-Esteem		-110.910						
		(62.088)						
Target × Self-Esteem		0.041						
		(0.061)						
Target^2 × Self-Esteem		-0.000						
		(0.000)						
Risk Attitude			-114.567					
			(69.808)					
Target × Risk Attitude			0.108					
			(0.066)					
Target^2 × Risk Attitude			-0.000					
			(0.000)					
Loss Aversion				149.873				
				(65.812)				
Target × Loss Aversion				-0.118				
				(0.064)				
Target^2 × Loss Aversion				0.000				
				(0.000)				
Intellect Imagination					45.900			
					(69.245)			
Target × Intellect Imagination					-0.026			
					(0.066)			
Target^2 × Intellect Imagination					0.000			
					(0.000)			
Gender: Female						-589.159		
						(453.131)		
Gender: Male						-755.579		
						(456.149)		
Target × Gender: Female						-0.044		
						(0.468)		
Target^2 × Gender: Female						0.000		
						(0.000)		
Target × Gender: Male						0.246		
						(0.471)		
Target^2 × Gender: Male						0.000		
						(0.000)		
Age							1.526	
							(4.285)	
Target × Age							-0.011	
							(0.004)	
Target^2 × Age							0.000	
							(0.000)	
Task Self-Eff. Q2								8.614
								(173.446)
Task Self-Eff. Q3								18.801
								(169.259)
Task Self-Eff. Q4								218.515
								(228.562)
Task Self-Eff. Q5								-2.854
								(228.115)
Target × Task Self-Eff. Q2								-0.007
								(0.176)
Target^2 × Task Self-Eff. Q2								0.000
								(0.000)
Target × Task Self-Eff. Q3								0.147
								(0.167)
Target^2 × Task Self-Eff. Q3								-0.000
								(0.000)
Target × Task Self-Eff. Q4								0.142
								(0.221)
Target^2 × Task Self-Eff. Q4								-0.000
								(0.000)
Target × Task Self-Eff. Q5								0.349
								(0.217)
Target^2 × Task Self-Eff. Q5								-0.000
								(0.000)
Observations	2804	2804	2804	2804	2804	2804	2804	2804
R-squared	0.057	0.058	0.056	0.055	0.054	0.066	0.079	0.118

This table reports the same OLS specifications as Tables 7 and 8. Reported p-values are adjusted for multiple hypothesis testing using the bootstrap step-down procedure of List et al. [2019], [2023], controlling the family-wise error rate across all estimated models. Robust standard errors are reported in parentheses. Statistical significance is denoted by $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

B Appendix: Experimental Screens

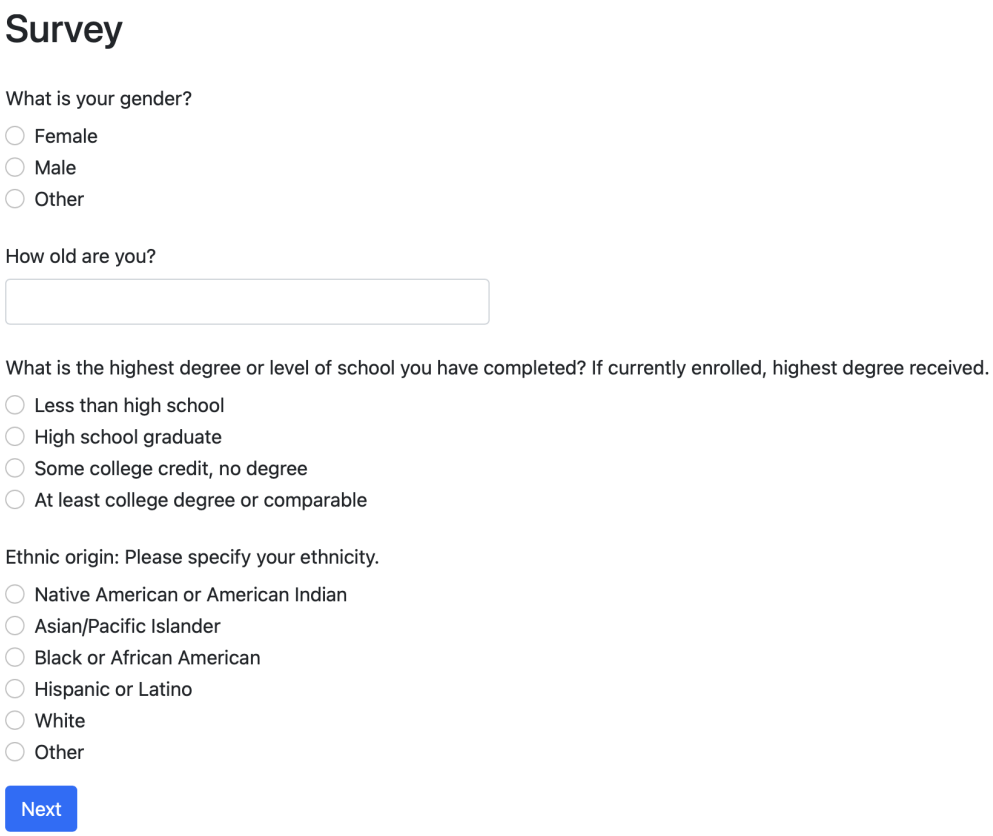

Figure 11: Demographics of the Participants (Survey)

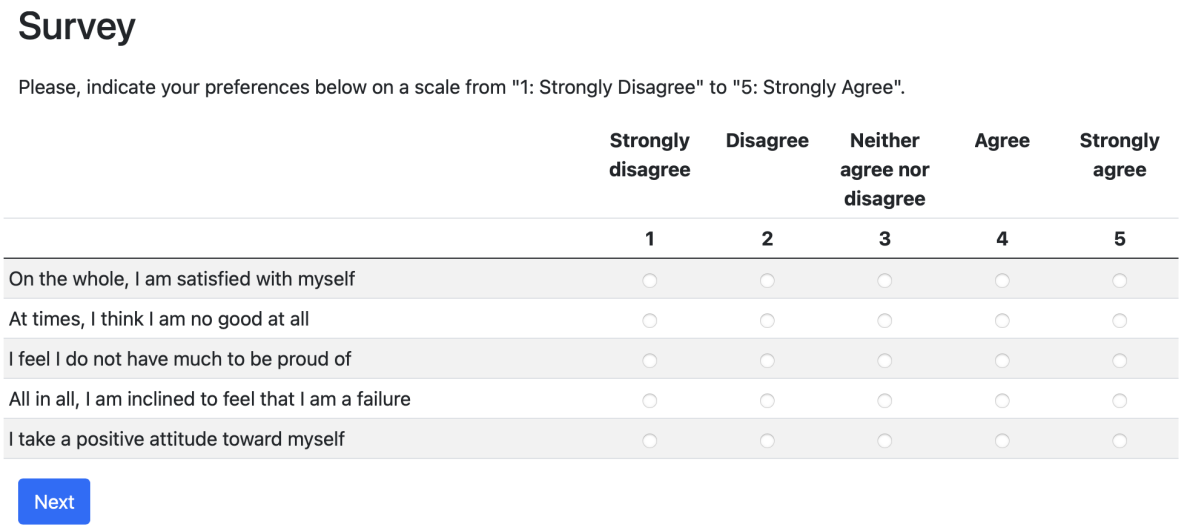

Figure 12: Survey Questions on Self-Esteem

Survey

Please, indicate your preferences below on a scale from "0: I do not want to take risks" to "10: I like to take risks".

To what extent do you like to take risks in ...

	I do not want to take risks					I am neutral about taking risks					I like to take risks
	0	1	2	3	4	5	6	7	8	9	10
... general	○	○	○	○	○	○	○	○	○	○	○
... financial matters	○	○	○	○	○	○	○	○	○	○	○
... career decisions	○	○	○	○	○	○	○	○	○	○	○

Please, indicate your preferences below on a scale from "1: Strongly Disagree" to "5: Strongly Agree".

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
	1	2	3	4	5
I am ambitious	○	○	○	○	○
I strive for success	○	○	○	○	○
I have challenging goals	○	○	○	○	○
For me it is very important to achieve outstanding results in my life	○	○	○	○	○
For me it is very important to accomplish great things	○	○	○	○	○

Next

Figure 13: Survey Questions on Risk Tolerance and Ambitiousness

Survey

Please, indicate your preferences below on a scale from "1: Strongly Disagree" to "5: Strongly Agree".

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
	1	2	3	4	5
When making a decision, I think much more about what might be lost than what might be gained.	○	○	○	○	○
The pain of losing money matters more than the pleasure of gaining the same amount of money.	○	○	○	○	○
I feel nervous when I have to make a decision that may lead to loss.	○	○	○	○	○
The pain from losing something matters much more to me than the pleasure from getting it.	○	○	○	○	○
Experiencing a major loss stays in my mind longer than experiencing a major gain.	○	○	○	○	○
A potential failure scares me more than a potential success encourages me.	○	○	○	○	○
The suffering that comes with losses can be fully offset by the pleasure that comes from gains.	○	○	○	○	○

Next

Figure 14: Survey Questions on Loss Aversion

Survey

Here are a number of personality traits that may or may not apply to you. Please select a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
I can tell about myself that I:	**1**	**2**	**3**	**4**	**5**
Am the life of the party.	○	○	○	○	○
Sympathize with others' feelings.	○	○	○	○	○
Get chores done right away.	○	○	○	○	○
Have frequent mood swings.	○	○	○	○	○
Have a vivid imagination.	○	○	○	○	○
Don't talk a lot.	○	○	○	○	○
Am not interested in other people's problems.	○	○	○	○	○
Often forget to put things back in their proper place.	○	○	○	○	○
Am relaxed most of the time.	○	○	○	○	○
Am not interested in abstract ideas.	○	○	○	○	○
Talk to a lot of different people at parties.	○	○	○	○	○
Feel others' emotions.	○	○	○	○	○
Like order.	○	○	○	○	○
Get upset easily.	○	○	○	○	○
Have difficulty understanding abstract ideas.	○	○	○	○	○
Keep in the background	○	○	○	○	○
Am not really interested in others.	○	○	○	○	○
Make a mess of things.	○	○	○	○	○
Seldom feel blue.	○	○	○	○	○
Do not have a good imagination.	○	○	○	○	○

Figure 15: Survey Questions on Big-5 Personality Traits

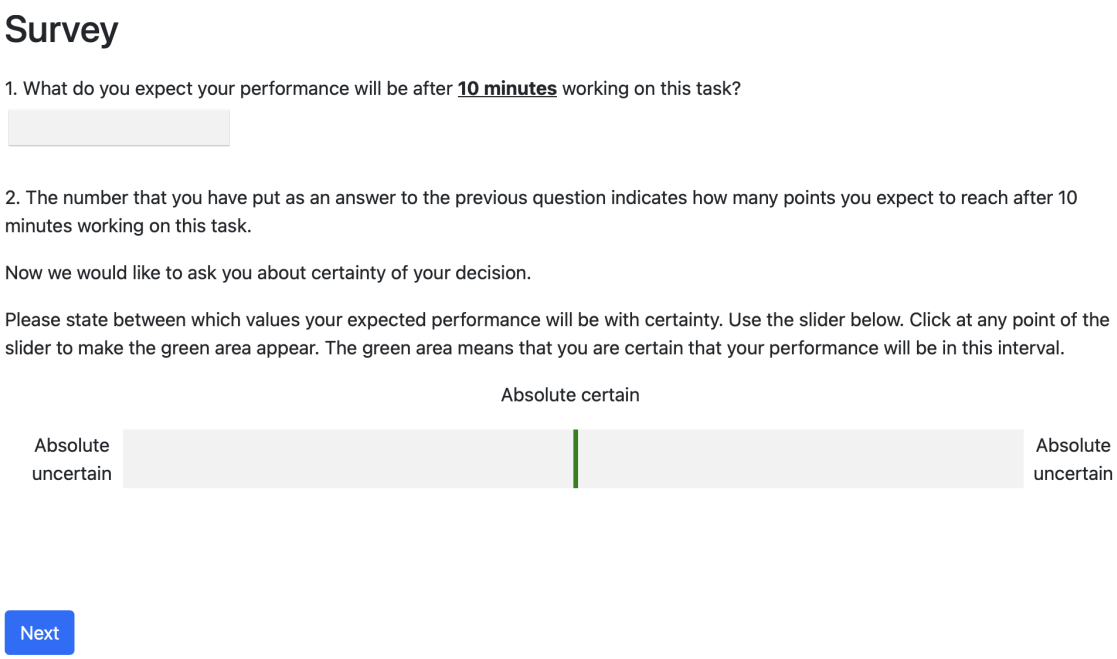

Figure 16: Survey Questions on Task Self-Efficacy

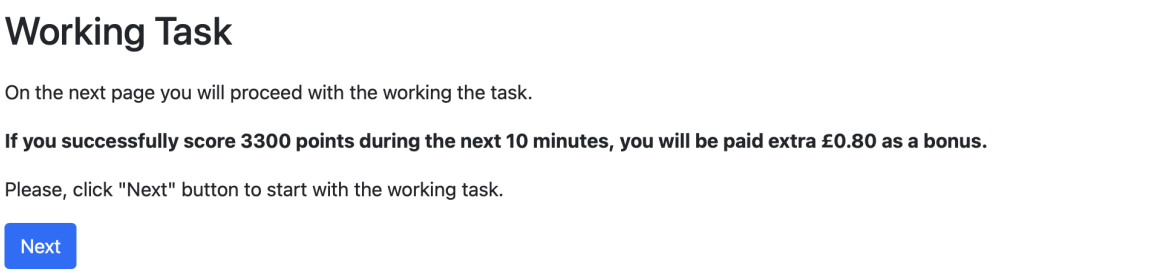

Figure 17: Target Assignment

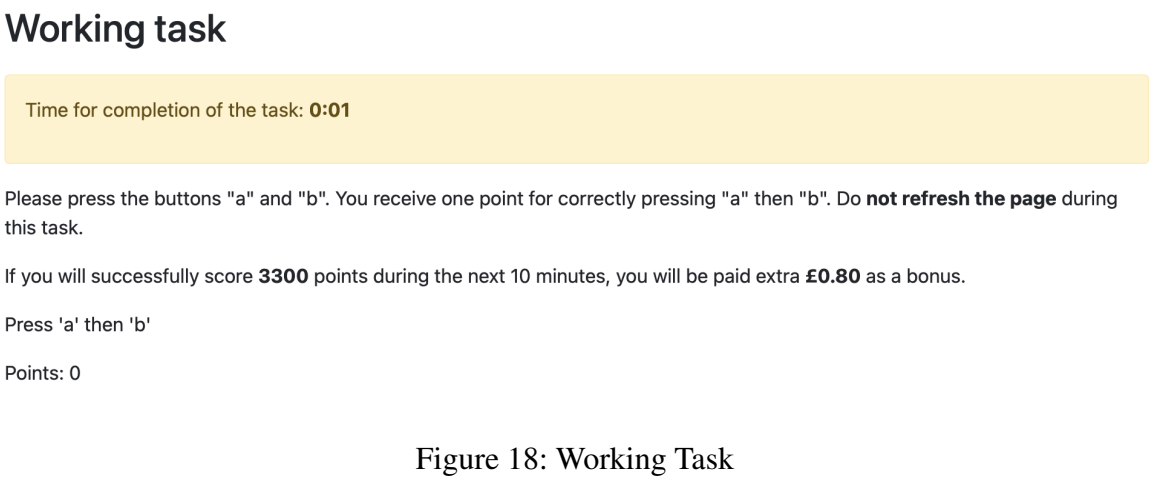

Figure 18: Working Task

Survey

Please, indicate whether you agree with the statements below on a scale from "1: Strongly disagree" to "5: Strongly agree".

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
	1	2	3	4	5
I enjoyed working on the task.	○	○	○	○	○
I felt stressed while completing the task.	○	○	○	○	○
I would like to participate in another task of this kind.	○	○	○	○	○

Next

Figure 19: Survey Questions on Satisfaction

Results

- You have successfully gained **0** points in the second part of the task.
- You have not reached **3300** points, and therefore you do not receive an additional bonus.
- Your final payment is **£1.70**.

Next

Figure 20: Example of Results and Payoffs