Bringing it all back home: Incentives in the age of general population sampling

MORITZ JANAS
New York University Abu Dhabi, Center for Behavioral Institutional Design, e-mail: moritz.janas@nyu.edu

LINA LOZANO
New York University Abu Dhabi, Center for Behavioral Institutional Design, e-mail: lina.lozano@nyu.edu

NIKOS NIKIFORAKIS
New York University Abu Dhabi, Center for Behavioral Institutional Design, e-mail: nikos.nikiforakis@nyu.edu

ERNESTO REUBEN
New York University Abu Dhabi, Center for Behavioral Institutional Design, e-mail: ereuben@nyu.edu

ROBERT STÜBER
University of Warwick, Center for Behavioral Institutional Design, e-mail: robert.stueber@wbs.ac.uk

ABSTRACT

Monetary incentives have long been a cornerstone of economic experiments. However, unincentivized measures of economic preferences and skills are becoming more common as experimenters move beyond the lab to study general population samples. This paper examines how monetary incentives influence inferences about truth-telling, competitiveness, and cognitive skills using a large, nationally representative U.S. sample. We find that incentives substantially alter the levels and patterns of truth-telling, competitiveness, and cognitive skills and increase the time participants spend reading instructions and making decisions. Crucially, in numerous instances, monetary incentives affect the conclusions derived from the data concerning group differences (e.g., age groups, gender, income groups, and educational attainment) as well as the estimated associations between income and the measured preferences/skills.

This version: February 2025

ACKNOWLEDGEMENTS

We thank Gabriel Møller for his research assistance. We gratefully acknowledge financial support from Tamkeen under the NYU Abu Dhabi Research Institute Award CG005. The project was approved by the NYU Abu Dhabi Institutional Review Board (HRPP-2022-110).

> "Although 'asking purely hypothetical questions is inexpensive, fast and convenient' (Thaler, 1991, p. 120), we conjecture that the benefits of being able to run many studies do not outweigh the costs of generating results of questionable reliability"—Hertwig and Ortmann (2001)

1. Introduction

Monetary incentives have long been a cornerstone of economic experiments (Smith, 1982). Whether used to induce preferences for testing economic theories (Smith, 1976) or to align the choices of participants with their true preferences (Camerer and Hogarth, 1999; Grether and Plott, 1979), economists have always considered incentivized tasks fundamental to ensuring the validity of experimental findings (Falk and Heckman, 2009). This conviction is perhaps best encapsulated in the *Aims & Scope* of *Experimental Economics*, which states: "[W]e only consider studies that do not employ deception of participants and in which participants are incentivized." As Hertwig and Ortmann put it, "experimental economists who do not use [monetary incentives] at all can count on not getting their results published" (Hertwig and Ortmann, 2001, p. 390). But that was then, when experimental economics research was almost exclusively conducted in purpose-built laboratories on university campuses with student samples (Nikiforakis and Slonim, 2019).

As experimental economists began to venture beyond the confines of the laboratory to assess the external validity of their findings in large, general population samples, the rising cost of monetary incentives led some to adopt hypothetical tasks (Falk et al., 2018, 2023). The availability of large, heterogeneous samples allowed researchers to investigate novel and ambitious questions, leading to publications in top academic journals (Falk et al., 2018; Falk and Hermle, 2018; Falk et al., 2021). As a result, hypothetical tasks gradually shed their taboo status, and an increasing number of experimenters began using them in general population studies (e.g., Bokern et al., 2023; Buser et al., 2025; Hauge et al., 2023). However, this methodological shift raises a fundamental question: to what extent does the absence of incentives influence the conclusions drawn from these large-scale datasets? This concern is particularly relevant when comparing preferences across demographic groups—a key objective of many hypothetical-choice experiments—where incentives may differentially affect participants' behavior based on their socioeconomic background (Bühren and Kundt, 2015; Falk et al., 2021; Gneezy et al., 2019).[1]

In this paper, we address this question by conducting a study with a large, nationally representative sample of the U.S. population. We vary the use of monetary incentives to study how they affect the measurement of economic *preferences*—specifically, competition and truth-telling—as well as *skills*. Measurement of cognitive skills has long been the domain of psychology, where experimenters do not require salient monetary incentives.[2] In fact, cognitive psychologists argue that monetary incentives should *not* be used when measuring skills because they can crowd out intrinsic motivation (Deci, 1971; Lepper and Greene, 2015).[3] While psychology's focus on skills may explain some of the disciplinary differences in attitudes toward unincentivized tasks, some economists argue that monetary incentives can improve attention in general (Bronchetti et al., 2023) and allow them specifically to obtain an accurate measure of cognitive skills by minimizing the influence of motivation (Alaoui and Penta, 2022; Proto et al., 2022). Given the growing interest of experimental economists in measuring cognitive skills (Alaoui and Penta, 2022; Falk et al., 2021; Proto et al., 2019, 2022), it is essential to understand how incentives affect their measurement.

Laboratory experiments with student samples have provided evidence that incentives affect the extent of truth-telling (Fischbacher and Föllmi-Heusi, 2013). So much so, that Charness et al. (2019) find no evidence of cheating in a die-roll task when monetary incentives are absent. Incentives have also been shown to enhance the performance of students in IQ tests and problem-solving activities (Gneezy and Rustichini, 2000; Gneezy et al., 2019).[4] Our study extends this research not only by employing a non-student sample but also by investigating how incentives affect inferences made regarding group differences. This is important for two reasons: (*i*) group comparisons have been a focus of the above-mentioned experimental studies (Falk et al., 2018; Falk and Hermle, 2018; Falk et al., 2021); (*ii*) evidence indicates that different groups can sometimes react differently to incentives (Gneezy et al., 2019; Sittenthaler and Mohnen,

[1] Hertwig and Ortmann (2001) review empirical evidence on how incentives influence decisions in laboratory experiments with student samples.

[2] As Roth (1995) remarked, "the question of actual versus hypothetical choices has become one of the fault lines that have come to distinguish experiments published in economic journals from those published in psychology journals" (p. 86).

[3] Indeed, one of the classic measures of cognitive skills—Raven's matrices—relies on unincentivized tasks (Raven, 1936, 2003).

[4] Gneezy and Rustichini (2000) compare performance in an IQ test without monetary incentives, with a fixed payment, or one of two performance-based schemes—with high and low incentives per correct answer. They find that high incentives improve performance, but low incentives can negatively affect performance.

2020). We consider four group comparisons that are commonly studied by economists: (*i*) male and female participants, (*ii*) high- and low-income participants, (*iii*) participants with and without a college degree, and (*iv*) younger and older participants. To the best of our knowledge, our experiment is the first to test how incentives affect individuals' willingness to compete.[5] Evidence also suggests that incentives can enhance both the accuracy and precision of behavioral measures in lab experiments. For instance, Burke et al. (1996) and Harrison (1994) find fewer violations of expected utility theory when monetary incentives are provided. We are unaware of any study on how incentives influence the time participants spend reading experimental instructions and making economic decisions in an online experiment. The time spent reading instructions could be considered an indicator of the quality of participants' choices.

Our main finding is that incentives critically affect the conclusions drawn from the data concerning group differences, both quantitatively and qualitatively. In half of the instances (6 out of 12), we would draw different conclusions concerning group differences from samples facing hypothetical tasks and samples with salient monetary incentives. Moreover, we find that using incentives changes the relationship between elicited preferences and income. Specifically, only incentivized measures of competitiveness and dishonesty exhibit significant correlations with income. In contrast, both cognitive skill measures are correlated with income, but the association is stronger in the unincentivized task. In addition, we show that monetary incentives increase the time spent by participants reading instructions, increase the likelihood of lying, and improve performance on Raven's matrices. By contrast, they reduce the share of individuals who choose to compete.

The rest of the paper is organized as follows. Section 2 describes the study design. Section 3 presents the results. Finally, Section 4 concludes by discussing our findings, mechanisms, and implications.

2. Experimental design

2.1. Overview

We investigate how incentives affect the measurement of competitiveness, truth-telling, and cognitive skills. All three measures have received substantial attention in the experimental

[5] Buser et al. (2025) collected data using the classic willingness-to-compete design of Niederle and Vesterlund (2007), with and without monetary incentives. However, the focus of their study is different, and the authors do not analyze how incentives affect participants' willingness to compete.

economics literature. In addition, they have been found to correlate with real-world behaviors/outcomes.[6] The experiment consists of two conditions that differ in whether participants' choices are incentivized (treatment *Incentivized*) or not (treatment *Hypothetical*). We also collect meta-data concerning the time a participant spent on each screen of the experiment—both instruction and decision screens. We begin by describing the design for treatment *Incentivized* and proceed to discuss treatment *Hypothetical*.

2.2. *Incentivized* treatment

As is common in experiments using incentivized tasks with large samples, we use a probabilistic payment scheme (Aydogan et al., 2024; Charness et al., 2016). Specifically, we randomly select 10% of participants for payment (200 out of the 2,000). Participants are also informed that the study consists of several parts and that the computer will randomly determine one part to be used for payment.

2.2.1. Measuring competitiveness

We measure competitiveness using a variation of the experimental design of Niederle and Vesterlund (2007). Participants choose the incentive scheme for a real-effort task that consists of correctly counting the number of 1's in 4 x 4 tables consisting of 1's and 0's for 45 seconds. Participants view one table at a time and do not receive feedback about whether their answer is right or wrong. Participants perform the task once.[7]

Before performing the task, participants must choose one of two payment schemes. If a participant chooses *Not compete*, they earn \$1 for every correct answer. If a participant chooses *Compete*, their number of correct answers is compared to the number of correct answers of another individual selected randomly from the sample. If the participant has more correct answers, they earn \$2; otherwise, they earn nothing. The choice of payment scheme is our measure of competitiveness.[8]

[6]Evidence shows that competitiveness predicts life outcomes such as income, educational attainment, and early motherhood (Buser et al., 2014, 2017, 2022; Dariel and Nikiforakis, 2022; Dariel et al., 2024; Reuben et al., 2017, 2024). Potters and Stoop (2016) show that dishonesty in the Mind Game predicts dishonesty outside the lab. Finally, cognitive skills are correlated with a variety of labor market outcomes (Heckman and Vytlacil, 2001; Heckman et al., 2006).

[7]Participants correctly counted the number of 1's in 4.2 tables on average.

[8]Given the design differences to Niederle and Vesterlund (2007), we conducted a pilot study with 1,500 participants to evaluate how they affect behavior. We observe similar gender differences in the willingness to select into competition across designs ($p = 0.692$ for the difference-in-differences estimate).

2.2.2. Measuring truth-telling

To measure truth-telling, we use a version of the *Mind Game* (Jiang, 2013; Potters and Stoop, 2016; Shalvi and De Dreu, 2014). Participants are asked to think of an integer number between 1 and 10. We then show them a randomly drawn number $U \sim [1, 10]$. Participants are subsequently asked to report whether the number they were shown matched the number they thought of. If a participant indicates that the two numbers match, their bonus earnings are $20. Otherwise, if a participant reports that the numbers do not match, their bonus earnings are only $10.

The *Mind Game* allows us to determine the extent of truth-telling and dishonesty at the group level. If all participants are honest, the share reporting that the numbers matched would be 10%. Hence, if we assume that few participants report that the numbers did not match even when they matched and observe that $(10+x)\%$ of individuals in a given group reported matching numbers, we can infer that $x\%$ were dishonest. In our analysis below, we will focus on the share of participants reporting that the numbers matched the randomly drawn number.

2.2.3. Measuring cognitive skills

We measure participants' cognitive skills using the Raven's Progressive Matrices test (Raven, 1936, 2003)—a non-verbal test often used by psychologists to measure general human intelligence and abstract reasoning skills, and more recently also by economists (e.g., Gill and Prowse, 2016; Proto et al., 2019; Proto et al., 2022). A Raven's Progressive Matrix consists of an image containing abstract geometric patterns following a logical progression. A part of the image is omitted, and participants must correctly identify the missing piece from a set of options provided to them.

Following Bilker et al. (2012) and Mani et al. (2013), we do not impose any time restriction. We implement a condensed version of the test, consisting of nine matrices proposed and validated by Bilker et al. (2012). Participants earn $2 for each correctly solved matrix and submit their answer for one matrix at a time. Like with the competitiveness measure, participants do not receive feedback about their performance at any time.

2.3. *Hypothetical* treatment

To facilitate a clean test of the role of monetary incentives, we aimed to keep the experimental instructions as similar as possible across treatments. For this reason, we kept the discussion of monetary payments unchanged. However, in the *Hypothetical* treatment, we added the following statement at the start of the experimental instructions:

> *"In some parts, you will be asked to make choices to earn money as a "bonus payment". Please note that while monetary amounts are presented as a currency, these amounts are entirely hypothetical and for the purpose of this study only. You will not be paid the bonus payment. However, even though the monetary amounts are hypothetical, please make your choices as if they are real."*

In the case of cognitive skills, to align our design with how cognitive skills are typically measured by psychologists, rather than discussing hypothetical incentives, we do not mention bonus payments when describing the Raven's Matrices test. The experimental instructions for both treatments are provided in Appendix B.

2.4. A nationally representative sample of the U.S. population

Data collection occurred between August 14, 2023, and June 26, 2024. The experiment was administered to a sample of 3,000 individuals drawn from the U.S. population. Of them, 2,000 individuals were assigned to the *Incentivized* and 1,000 to the *Hypothetical* treatment.[9] Specifically, we determined nested quotas for gender and education (2×2), as well as quotas for age (four categories between 25 and 65). The age quotas were obtained from the 2022 UN World Population Prospects database (United Nations, 2022), while the gender/education quotas were taken from the UNESCO Institute for Statistics (2022).

Recruitment was undertaken by the market research company Ipsos using their U.S. panel of participants. Our sampling strategy allowed for small deviations from the target quotas such that recruitment was completed within a reasonable time frame. We reweigh our data throughout the paper using the target quotas such that our sample is nationally representative of the U.S. population in terms of age, gender, and education. However, all results reported in the paper are unaffected if we use the unweighted sample.

[9] The data was collected as part of a larger data collection effort (see https://osf.io/jng2r/?view_only=74c704955107485a982941dd6033a018, for more details). This dataset focuses on incentivized choices. Hence, we oversampled individuals for the *Incentivized* treatment.

To the best of our knowledge, this is the first time these economic games have been played with a representative U.S. sample. While this is a contribution of our study, what is more important for our purposes is that the diverse sample enables us to perform a variety of group comparisons, which are of special interest given that hypothetical tasks in economics are primarily utilized in general population samples.

As mentioned, we compare the effect of incentives across four groups: (*i*) men and women, (*ii*) individuals with above- and below-median after-tax income, (*iii*) individuals with and without a college degree, and (*iv*) participants older or younger than the weighted sample median age of 43 years. Gender, age, income, and educational attainment are self-reported (see also Appendix B).

3. Results

3.1. Screen-time differences

Before analyzing how incentives influence participants' decisions, we first assess whether incentives affect the time spent on experimental screens, either reading instructions or making decisions. If participants respond to hypothetical incentives as they would to real monetary incentives, time spent on different screens should be similar across treatments. We test this relationship in Table 1 using linear regressions of the time participants spent on the various screens on a dummy variable indicating whether they were in the *Incentivized* treatment. Since instructions varied slightly in the cognitive skills task (see Section 2.3), we measure time spent reading instructions as seconds per word in these regressions.[10] All regressions report robust standard errors.

We find that monetary incentives significantly increase the time participants spend reading instructions in all three tasks ($p < 0.012$). The magnitude of the effect is substantial: on average, participants spend from 32.4% more time reading the Raven's matrices instructions to 49.1% more time reading the competitiveness task instructions. These findings align with prior research demonstrating that monetary incentives influence participants' attention during experiments (e.g., Harrison, 1994; Burke et al., 1996) and suggest incentives increase the time

[10] On average, participants spent 104 seconds on the competitiveness task—87 seconds reading the instructions and 17 seconds deciding whether to compete. In the truth-telling task, they spent 30 seconds reading the instructions and 10 seconds deciding whether to claim the numbers matched. Finally, in the Raven's matrices task, participants spent 62 seconds reading the instructions and 424 seconds completing the matrices.

Table 1. Time spent on instructions and tasks

	Choosing competition		Claiming numbers matched		Raven's matrices	
	Instructions (1)	Choice (2)	Instructions (3)	Choice (4)	Instructions (5)	Perform (6)
Incentivized	0.137^{***}	1.802	0.131^{***}	1.345^{***}	0.175^{**}	51.041^{***}
	(0.035)	(1.379)	(0.035)	(0.518)	(0.069)	(15.722)
Constant	0.288^{***}	16.110^{***}	0.267^{***}	9.283^{***}	0.543^{***}	387.228^{***}
	(0.012)	(0.833)	(0.008)	(0.231)	(0.039)	(12.120)
N	3000	3000	3000	3000	3000	3000

Notes: OLS regressions of the number of seconds spent per word on the instructions of the competitiveness task (1), truth-telling task (3), and Raven's matrices (5). OLS regressions of the number of seconds spent deciding whether to compete (2), choosing whether to claim the numbers matched (4), and performing the nine Raven's matrices (8). In all regressions, the dependent variable equals one for participants in the *Incentivized* treatment and zero for those in the *Hypothetical* treatment. Robust standard errors are reported in parentheses. Estimates are weighed to be nationally representative of the U.S. population in terms of age, gender, and education. ***, **, and * indicate statistical significance at 1%, 5%, and 10%.

they dedicate to understanding the instructions.

Incentives also increase participants' decision-making time. In the truth-telling task, participants in the *Incentivized* treatment took 1.35 seconds longer to decide whether to claim the numbers matched ($p = 0.009$), a 14.4% increase. In the competitiveness task, participants took 1.80 seconds longer in the *Incentivized* treatment (an 11.2% increase), though this difference is not statistically significant ($p = 0.191$). An interpretation for the longer response times is that participants perceive the trade-offs inherent in their choice as more difficult when it involves real instead of hypothetical incentives (e.g., see Alós-Ferrer et al., 2021). Finally, in the Raven's matrices task, participants in the *Incentivized* treatment spent 51.0 seconds longer solving the nine matrices, a 13.2% increase ($p = 0.001$). This finding supports recent work arguing that incentives enhance cognitive effort in reasoning tasks (Alaoui and Penta, 2022).[11]

To summarize, we find that incentives influence the time participants spend reading instructions and making decisions. Next, we examine whether they also influence the decisions they make.

3.2. Level differences

Figure 1 depicts the share of participants choosing to compete, the share reporting the numbers matched, and the mean number of correct Raven's matrices in the *Hypothetical* and *Incentivized*

[11] All these results are robust to (i) excluding 5% of outliers who spent the longest time on each screen and (ii) considering the natural logarithm of time as the dependent variable.

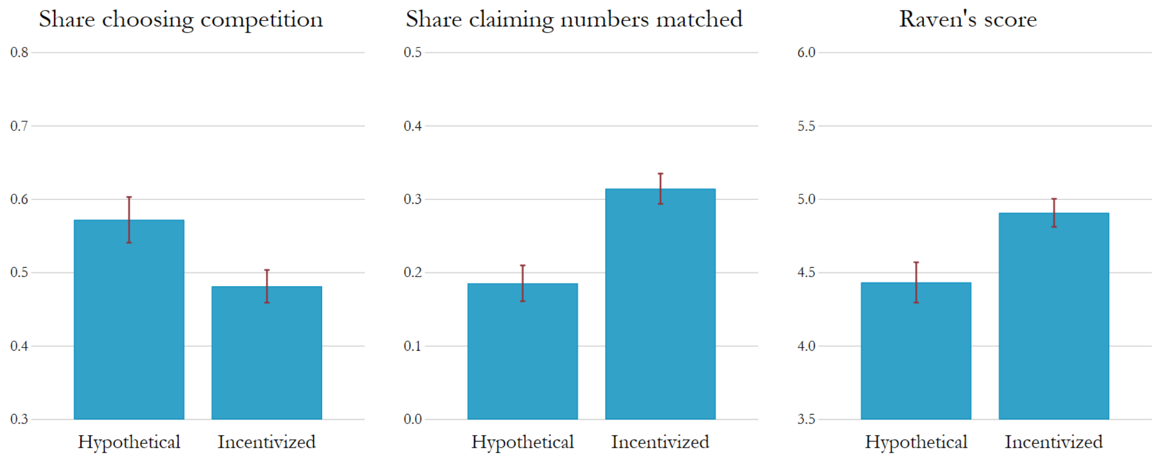


Figure 1. Share of participants choosing to compete (left), share claiming the numbers matched (middle), and mean number of correct Raven's matrices (right) by treatment.

Notes: Error bars correspond to 95% confidence intervals.

treatments.

The figure demonstrates there are notable treatment differences. First, the share of participants choosing to compete is 9.1 percentage points lower in the *Incentivized* treatment compared to the *Hypothetical* treatment (48.1% vs. 57.2%), a difference of 0.181 standard deviations. Second, the share of participants claiming the number they thought of matched the displayed number is 12.8 percentage points higher in the *Incentivized* than in the *Hypothetical* treatment (31.4% vs. 18.6%), a difference of 0.289 standard deviations. Finally, on average, participants solve 0.475 Raven's matrices more when they have real rather than hypothetical incentives (4.908 vs. 4.433 matrices), a difference of 0.219 standard deviations.

We test whether these treatment differences are statistically significant in Table 2. The table presents the results of linear regressions of choosing to compete in columns (1) to (3), claiming the numbers matched in columns (4) to (6), and the fraction of correctly-solved Raven's matrices in columns (7) to (9). Columns (1), (4), and (7) estimate the difference between the *Incentivized* and *Hypothetical* treatments without any controls, confirming that the differences depicted in Figure 1 are highly significant for all three outcomes ($p < 0.001$).[12] Since the *Incentivized* treatment introduces real financial stakes, participants' income could potentially

[12]The experimental literature on competitiveness often distinguishes between the choice to compete and the degree to which this choice is driven by participants' preferences (for risk or competition). A common approach to isolate the preference component of the choice to compete is to control for participants' performance and beliefs in the regression (e.g., Niederle and Vesterlund, 2007; Buser et al., 2014; Lozano and Reuben, 2022). Table A5 in the Appendix shows that while beliefs about relative performance significantly predict competition entry, the negative effect of incentives remains unchanged.

Table 2. Treatment differences in elicited behavior

	Choosing competition			Claiming numbers matched			Raven's score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Incentivized	-0.090^{***}	-0.092^{***}	-0.091^{***}	0.129^{***}	0.125^{***}	0.123^{***}	0.053^{***}	0.052^{***}	0.046^{***}
	(0.020)	(0.020)	(0.020)	(0.016)	(0.016)	(0.016)	(0.009)	(0.009)	(0.009)
Constant	0.572^{***}	0.573^{***}	0.573^{***}	0.186^{***}	0.188^{***}	0.190^{***}	0.493^{***}	0.493^{***}	0.498^{***}
	(0.016)	(0.016)	(0.016)	(0.012)	(0.013)	(0.013)	(0.008)	(0.008)	(0.008)
N	3000	3000	3000	3000	3000	3000	3000	3000	3000
Controls									
Income	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Time	No	No	Yes	No	No	Yes	No	No	Yes

Notes: OLS regressions of participants' choice to compete in columns (1) to (3), claim that the numbers matched in columns (4) to (6), and fraction of correctly-solved Raven's matrices in columns (7) to (9). In all regressions, the dependent variable equals one for participants in the *Incentivized* treatment and zero for those in the *Hypothetical* treatment. Income controls consist of 12 dummy variables, one for each income bin. Time controls consist of the standardized number of seconds participants took to read the instructions and complete the task. Robust standard errors are reported in parentheses. Estimates are weighed to be nationally representative of the U.S. population in terms of age, gender, and education. ***, **, and * indicate statistical significance at 1%, 5%, and 10%.

affect the impact of incentives on behavior. To account for this, columns (2), (5), and (8) control for income using a dummy variable for each income bin. Controlling for income does not have a meaningful effect on the estimated treatment differences. Finally, given that incentives increase the time spent reading instructions and making decisions, we further control for these variables in columns (3), (6), and (9). The impact of incentives remains large and statistically significant.[13]

In sum, using a large non-student sample, we confirm that incentives reduce truth-telling (Fischbacher and Föllmi-Heusi, 2013; Charness et al., 2019) and enhance performance in IQ tests (Gneezy and Rustichini, 2000; Gneezy et al., 2019). Additionally, we find that incentives decrease individuals' willingness to compete. Next, we examine whether incentives have heterogeneous effects across different groups and whether they impact the inferences we draw about group differences.

3.3. Group differences

To assess whether the effect of incentives varies across groups, Figure 1 shows the share of participants choosing to compete, the share reporting the numbers matched, and the mean

[13]Time spent solving Raven's matrices is positively correlated with the fraction of correct answers ($p < 0.001$). Moreover, more time spent making the truth-telling decision correlates with a higher likelihood of claiming the numbers matched ($p = 0.041$). Time spent is not significantly associated with competition entry.

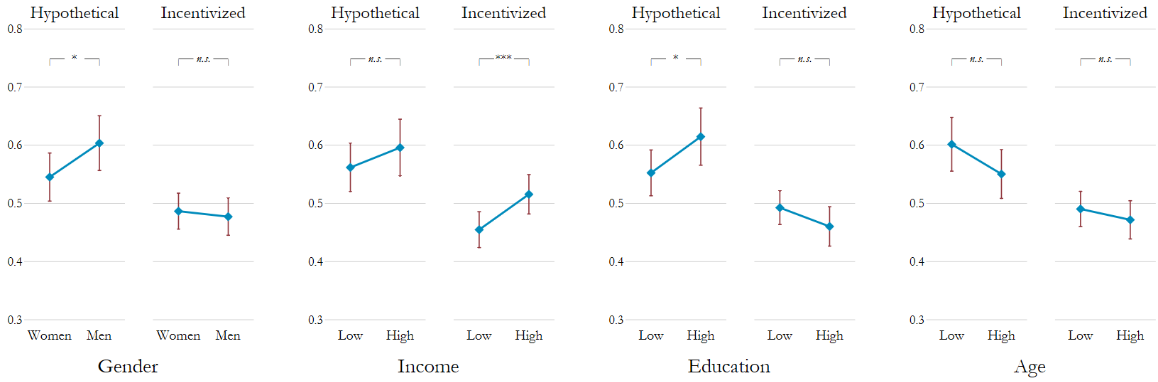
number of correct Raven's matrices, disaggregated by (*i*) gender (men vs. women), (*ii*) income (above vs. below the sample-median income), (*iii*) educational attainment (college vs. no college degree), and (*iv*) age (above vs. below the sample-median age), for each treatment.[14] For each group and treatment, we indicate whether the difference in behavior is statistically significant across the two group categories (***, **, and * indicate statistical significance at 1%, 5%, and 10%; *n.s.* means 'not significant' at the 10% level).

Figure 2 suggests that the effect of incentives varies across the different groups. Notably, this variation is sufficient that in 6 out of the 12 groups, we find a statistically significant group difference in one treatment but not in the other. In other words, the conclusions researchers draw might depend on whether they use monetary incentives or not.

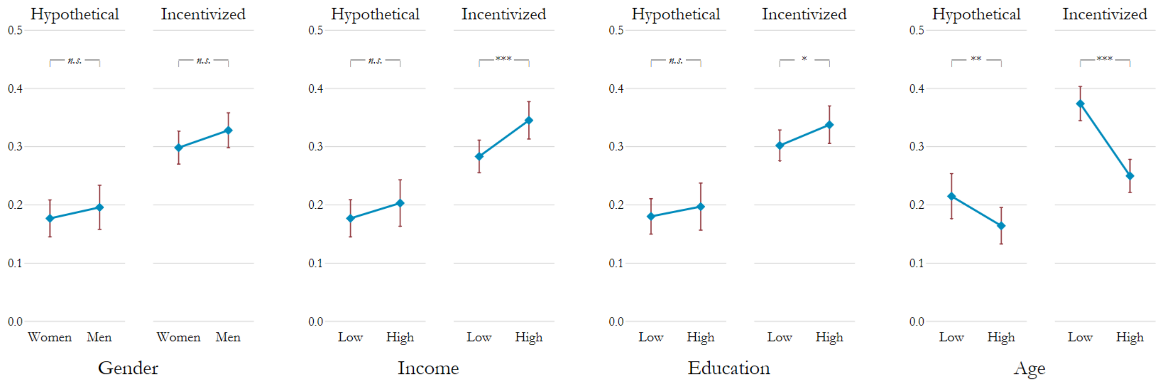
In Table 3, we test whether treatment differences vary significantly across groups. The table presents the results of linear regressions of choosing to compete in columns (1) to (4), claiming the numbers matched in columns (5) to (8), and the fraction of correctly-solved Raven's matrices in columns (9) to (12). We run a regression for each outcome and group combination. Each regression includes a treatment dummy, a group category dummy, and a treatment-group interaction as independent variables.

For competitiveness, we observe men choosing to compete more frequently than women in the *Hypothetical* treatment ($p = 0.068$). Incentives significantly reduce the share of competitive choices ($p < 0.026$), but the effect is stronger for men ($p = 0.084$), leading to a reversal of the gender gap in the *Incentivized* treatment ($p = 0.677$). A similar pattern arises for education: college-educated participants are more likely to compete in the *Hypothetical* treatment ($p = 0.053$), but their willingness to compete declines more sharply with incentives ($p = 0.015$), reversing the education gap ($p = 0.138$). We do not find differential responses to incentives for participants depending on their income or age. For truth-telling, Table 3 shows that the increase in participants claiming the numbers matched due to incentives does not vary significantly by gender, income, or education ($p > 0.290$). However, incentives have a smaller effect on truth-telling among older participants ($p = 0.026$). Lastly, for Raven's matrices, we find that the positive effect of incentives on performance is less pronounced for men ($p = 0.095$). Consequently, while men significantly outperform women in the *Hypothetical* treatment ($p = 0.006$), men and women perform equally well in the *Incentivized* treatment ($p = 0.307$). We observe a simi-

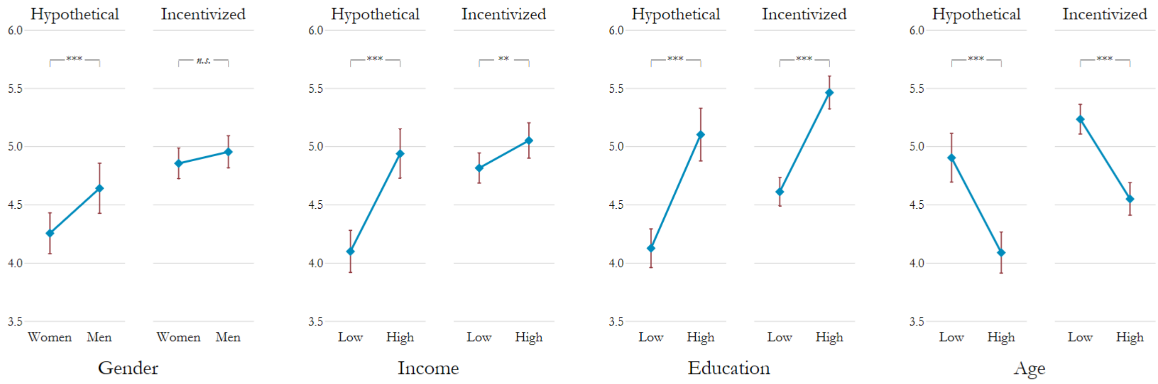
[14] We drop 141 participants who did not disclose their income from the income comparison and 3 participants who indicated their gender as 'other' from the gender comparison.



Panel A. Share of participants choosing to compete by treatment for various groups.



Panel B. Share of participants claiming the numbers matched by treatment for various groups.



Panel C. Mean number of correct Raven's matrices by treatment for various groups.

Figure 2. Treatment differences across different groups.

Notes: Error bars correspond to 95% confidence intervals. For each group and treatment, the figure indicates whether the difference in behavior is statistically significant across the two group categories. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, while *n.s.* indicates the difference is 'not significant,' i.e., $p > 0.10$.

Table 3. Treatment differences in elicited behavior for various groups

	Choosing competition				Claiming numbers matched				Raven's score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Incentivized	-0.059^{**}	-0.107^{***}	-0.059^{**}	-0.111^{***}	0.121^{***}	0.105^{***}	0.121^{***}	0.158^{***}	0.067^{***}	0.080^{***}	0.054^{***}	0.036^{***}
	(0.026)	(0.026)	(0.025)	(0.028)	(0.022)	(0.022)	(0.021)	(0.025)	(0.012)	(0.013)	(0.012)	(0.014)
Male	0.058^{*}				0.019				0.043^{***}			
	(0.032)				(0.025)				(0.016)			
Incentivized × Male	-0.068^{*}				0.011				-0.032^{*}			
	(0.039)				(0.033)				(0.019)			
High income		0.034				0.026				0.093^{***}		
		(0.033)				(0.026)				(0.016)		
Incentivized × High income		0.026				0.037				-0.067^{***}		
		(0.040)				(0.034)				(0.019)		
High education			0.062^{*}				0.017				0.108^{***}	
			(0.032)				(0.026)				(0.016)	
Incentivized × High education			-0.096^{**}				0.020				-0.014	
			(0.039)				(0.033)				(0.019)	
High age				-0.051				-0.051^{**}				-0.090^{***}
				(0.032)				(0.025)				(0.016)
Incentivized × High age				0.032				-0.073^{**}				0.015
				(0.039)				(0.033)				(0.019)
Constant	0.545^{***}	0.562^{***}	0.552^{***}	0.602^{***}	0.177^{***}	0.177^{***}	0.180^{***}	0.215^{***}	0.473^{***}	0.456^{***}	0.459^{***}	0.545^{***}
	(0.021)	(0.021)	(0.020)	(0.024)	(0.016)	(0.016)	(0.016)	(0.020)	(0.010)	(0.010)	(0.009)	(0.012)
N	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000

Notes: OLS regressions of participants' choice to compete in columns (1) to (4), claim that the numbers matched in columns (5) to (8), and fraction of correctly-solved Raven's matrices in columns (9) to (12). In all regressions, we include as dependent variables a treatment dummy for the *Incentivized* treatment, a group dummy identifying a group category, and a treatment-group interaction. The group categories are *Male* for the participants' gender, *High Income* for participants with income above the sample median, *High education* for participants with a college degree, and *High age* for participants above sample-median age. Robust standard errors are reported in parentheses. Estimates are weighed to be nationally representative of the U.S. population in terms of age, gender, and education. ***, **, and * indicate statistical significance at 1%, 5%, and 10%.

lar pattern for income. Without incentives, high-income participants outperform low-income participants by a substantial 0.387 standard deviations ($p < 0.001$). However, incentives lead to a larger increase in the scores of lower-income participants ($p = 0.001$), which considerably narrows the performance gap to a much smaller 0.109 standard deviations.[15]

3.4. Correlation with income

A common research strategy for studying the influence of preferences and skills is to measure them experimentally and correlate these measures with real-world outcomes (e.g., Levitt and List, 2007; Dohmen et al., 2010; Benjamin et al., 2013). One of the most relevant real-world outcomes for economists is income. It is, therefore, interesting to examine whether competitiveness, truth-telling, and cognitive skills are correlated with participants' income and whether these correlations depend on whether incentives are real or hypothetical. The results from this exercise can be found in Table 4. In all regressions, the dependent variable is the participants' self-reported after-tax monthly income (in thousands).[16] As independent variables, we use participants' choice to compete, claim that the numbers matched, and fraction of correct Raven's matrices.

In the *Hypothetical* treatment, neither choosing to compete ($p = 0.267$) nor claiming the numbers matched ($p = 0.112$) show a statistically significant correlation with income. By contrast, in the *Incentivized* treatment, the coefficients of both variables are substantially larger and are statistically significant. Specifically, participants who choose to compete report an average monthly income $328 (10.5%) higher than those who do not compete ($p = 0.004$), a result consistent with previous estimates (e.g., Buser et al., 2025; Reuben et al., 2024). Similarly, participants who claim the numbers matched report, on average, $470 (15.0%) higher income per month ($p < 0.001$).[17]

The correlation between cognitive skills and income also differs across treatments. In the *Hypothetical* treatment, the fraction of correct Raven's matrices is strongly correlated with income (coefficient: 1.784; $p < 0.001$), whereas in the *Incentivized* treatment, the correlation is

[15] Interestingly, similar patterns emerge when using the counting 1s task from the competitiveness measure as an alternative performance metric.

[16] Income was measured using 11 bins, ranging from 0 for participants without personal income to 11 for those earning more than $8,001 per month. We use the midpoint of each bin as the participant's income and assign $8,400 for those in the top bin.

[17] To our knowledge, no prior studies have examined the link between choices in the Mind Game and income. More broadly, evidence of the relationship between income (or socioeconomic status) and pro-social or immoral behavior is mixed (Andreoni et al., 2021).

Table 4. Regressions of income on competitiveness, truth-telling, and Raven's score

	Hypothetical			*Incentivized*		
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing competition	0.175			0.328^{***}		
	(0.161)			(0.115)		
Claiming numbers matched		0.342			0.453^{***}	
		(0.214)			(0.126)	
Raven's score			1.784^{***}			0.470^{*}
			(0.322)			(0.247)
Constant	3.070^{***}	3.107^{***}	2.292^{***}	3.135^{***}	3.152^{***}	3.035^{***}
	(0.122)	(0.087)	(0.170)	(0.079)	(0.068)	(0.146)
N	963	963	963	1896	1896	1896

Notes: OLS regressions of participants' self-reported after-tax monthly income. Income was measured using 11 bins, each consisting of a range of \$800. We use the midpoint of the selected bin as the participants' income. Choosing competition and claiming the numbers matched are binary variables, while Raven's score is the fraction of correct matrices (between 0 and 1). Robust standard errors are reported in parentheses. Estimates are weighed to be nationally representative of the U.S. population in terms of age, gender, and education. ***, **, and * indicate statistical significance at 1%, 5%, and 10%.

weaker and only marginally significant (coefficient: 0.470; $p = 0.070$). If we accept the argument by Gneezy et al. (2019) that unincentivized IQ tests capture both intrinsic motivation and cognitive ability, while incentivized IQ tests better isolate pure cognitive skills, then our results suggest that intrinsic motivation plays a crucial role in predicting income. Since much of the literature on IQ and labor market outcomes relies on unincentivized measures of cognitive skills (Heckman et al., 2006), our findings indicate that prior studies may overestimate the pure labor market returns to cognitive ability.

4. Discussion

Our study highlights the pivotal role of monetary incentives in measuring economic preferences and skills in general population samples. We find that incentives substantially alter both the levels and patterns of truth-telling, competitiveness, and cognitive skills, as well as the conclusions drawn about group differences. In half of the cases examined, group-level comparisons would lead to different inferences depending on whether tasks were incentivized or hypothetical. Moreover, incentives influence the relationship between economic preferences and income—only incentivized measures of competitiveness and dishonesty show significant correlations with income, while the association between cognitive skills and income is stronger in the unincentivized condition. Additionally, we find that incentives lead participants to engage more deeply with tasks, increasing time spent reading instructions and making decisions.

One question we have yet to address is why monetary incentives produce the effects observed in our experiment. In other words, what are the underlying mechanisms at play? Let us start with truth-telling. One plausible explanation for the effect of incentives on the levels of dishonesty observed in our experiment involves image concerns. Participants may want to signal adherence to honesty norms—either to themselves (e.g., Bénabou and Tirole, 2006; Mazar et al., 2008) or to the experimenter (e.g., Zizzo, 2010). In the absence of monetary incentives, individuals might prefer to report that the numbers did not match to maintain a positive self-image or avoid perceived scrutiny. However, when financial rewards are sufficiently high to offset the psychological cost of appearing or being dishonest, participants may prioritize monetary gains over reputational concerns, leading to higher rates of dishonesty under the incentivized condition. A similar mechanism could explain our findings on competitiveness. When measuring competitiveness through participants' choices regarding incentive schemes in a real-effort task, we find that making the choice payoff-relevant reduces the proportion of participants opting for competition. This decline may indicate that some individuals prefer to appear competitive in hypothetical settings but adjust their behavior when real financial stakes are introduced.

Image concerns threaten both the internal and external validity of experimental findings (e.g., De Quidt et al., 2018) and are thus often cited as a justification for using monetary incentives. How can image concerns explain the different inferences drawn regarding group comparisons in our data? The answer is that image concerns can vary across groups of individuals. For instance, if men care more than women to be perceived as competitive, incentives can differentially impact their choices, as seen in Figure 2. Similarly, the fact that wealthier individuals lie more than their less well-off counterparts in the presence of incentives could be due to wealthier individuals caring less about their image in this task.

Image concerns could also help explain our findings concerning cognitive skills. The psychological literature suggests that monetary incentives may be unnecessary (even undesirable) if individuals are intrinsically motivated to perform well. Image concerns may at least partly explain where individuals derive their intrinsic motivation from. Unlike with competitiveness and truth-telling, introducing incentives does not create tension with image concerns. Nevertheless, image concerns could vary across individuals and affect our conclusions concerning group differences. For instance, one explanation why women improve their performance more than men when incentives are introduced is that women care less about how they are perceived on account of their performance in the cognitive task. Of course, intrinsic motivation could arise from how interesting a task is to individuals. This, too, could vary across groups, leading to biased estimates, and could explain why we observe greater performance improvements among lower-income participants, suggesting that wealthier individuals may be more intrinsically motivated to excel in such tasks even without financial incentives. Further research is needed to help us distinguish between the different drivers of choices in the absence of monetary incentives.

Our findings support the Hertwig-Ortmann conjecture featured at the outset of this paper. While unincentivized choices reduce the cost of studying large and highly diverse samples, questions remain about the extent to which they capture meaningful insights. This concern also extends to *experimentally validated surveys*, such as those used by Falk et al. (2018) and Falk and Hermle (2018). Experimental validation relies on using small(er) samples to identify combinations of survey questions and hypothetical choices that best predict behavior in incentivized tasks (Falk et al., 2023; Fallucchi et al., 2020). Our finding that different groups respond differently to incentives implies that for experimental validation to live up to its promise of "leverag[ing] the strengths of both experimental and survey approaches" one must recruit samples for the validation that are as diverse as those in the population of interest. Given the large number of hypothetical tasks, survey questions, and incentivized choices involved in these validation exercises, ensuring reliable experimental validation may be especially challenging when studying highly heterogeneous populations across different countries (see e.g., Kosfeld and Sharafi, 2024; Kosfeld et al., 2025).

We strongly support the goal of expanding research to diverse general population samples—this remains an important frontier for experimenters in both economics and psychology (Henrich et al., 2010). However, our findings suggest that achieving this goal should not come at the expense of the incentive structures that have long defined the field as incentives play a crucial role in shaping the conclusions drawn from such samples. As internet access continues to expand and technological advancements drive down the cost of online experiments (Oberlo, 2024), incentivized online studies provide a promising path forward for researchers seeking to study diverse populations at scale. That said, questions remain about designing effective incentives and study instruments for general population samples, as these samples are significantly more heterogeneous than traditional student populations—not only in terms of income but also in their ability to comprehend complex, incentivized tasks. Hence, a combination of incentivized and general survey questions such as those used to measure attitudes towards risk (Dohmen et al., 2011) or competitiveness (Buser et al., 2025) may help us understand behavioral diversity and its origins. While more research is needed to refine these methods, one conclusion seems clear: incentives remain essential in the age of general population sampling.

References

Alaoui, L. and Penta, A. (2022). Cost-benefit analysis in reasoning. *Journal of Political Economy*, 130(4):881–925.

Alós-Ferrer, C., Fehr, E., and Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129(6):1828–1877.

Andreoni, J., Nikiforakis, N., and Stoop, J. (2021). Higher socioeconomic status does not predict decreased prosocial behavior in a field experiment. *Nature Communications*, 12(1):4266.

Aydogan, I., Berger, L., and Théroude, V. (2024). Pay all subjects or pay only some? An experiment on decision-making under risk and ambiguity. *Journal of Economic Psychology*, 104:102757.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is 'behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11(6):1231–1255.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., and Gur, R. C. (2012). Development of abbreviated nine-item forms of the raven's standard progressive matrices test. *Assessment*, 19(3):354–369.

Bokern, P., Linde, J., Riedl, A., Schmeets, H., and Werner, P. (2023). The convergent and external validity of risk preference elicitation methods: Controlling for measurement error in a large population sample. Netspar Discussion Paper 08/2023-038.

Bronchetti, E. T., Kessler, J. B., Magenheim, E. B., Taubinsky, D., and Zwick, E. (2023). Is attention produced optimally? Theory and evidence from experiments with bandwidth enhancements. *Econometrica*, 91(2):669–707.

Bühren, C. and Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment and Decision Making*, 10(2):185–190.

Burke, M. S., Carter, J. R., Gominiak, R. D., and Ohl, D. F. (1996). An experimental note on the allais paradox and monetary incentives. *Empirical Economics*, 21:617–632.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Buser, T., Niederle, M., and Oosterbeek, H. (2025). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. *The Review of Economics and Statistics*, pages 1–45.

Buser, T., Peter, N., and Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American Economic Review: Papers & Proceedings*, 107(5):125–30.

Buser, T., Peter, N., and Wolter, S. C. (2022). Willingness to compete, gender and career choices along the whole ability distribution. *Experimental Economics*, 25(5):1299–1326.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19:7–42.

Charness, G., Blanco-Jimenez, C., Ezquerra, L., and Rodriguez-Lara, I. (2019). Cheating, incentives, and money manipulation. *Experimental Economics*, 22:155–177.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Dariel, A., Ham, J. C., Nikiforakis, N., and Stoop, J. (2024). Disparities in psychological traits and income: Race, ethnicity, and gender in the U.S. IZA Discussion Paper No. 16818.

Dariel, A. and Nikiforakis, N. (2022). Is there a motherhood gap in the willingness to compete for pay?: Evidence from the Netherlands, the UAE and the USA. NYUAD Working Paper No. 20220079.

De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1):105.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

Falk, A. and Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412):eaas9899.

Falk, A., Kosse, F., Pinger, P., Schildberg-Hörisch, H., and Deckers, T. (2021). Socioeconomic status and inequalities in children's iq and economic preferences. *Journal of Political Economy*, 129(9):2504–2545.

Fallucchi, F., Nosenzo, D., and Reuben, E. (2020). Measuring preferences for competition with experimentally-validated survey questions. *Journal of Economic Behavior & Organization*, 178:402–423.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124(6):1619–1676.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., and Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, 1(3):291–308.

Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4):623–638.

Harrison, G. W. (1994). *Expected utility theory and the experimentalists*. Springer.

Hauge, K. E., Kotsadam, A., and Riege, A. (2023). Culture and gender differences in willingness to compete. *The Economic Journal*, 133(654):2403–2426.

Heckman, J. and Vytlacil, E. (2001). Identifying the role of cognitive ability in explaining the level of and change in the return to schooling. *Review of Economics and Statistics*, 83(1):1–12.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83.

Hertwig, R. and Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(3):383–403.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, 93:328–336.

Kosfeld, M. and Sharafi, Z. (2024). The preference survey module: evidence on social preferences from tehran. *Journal of the Economic Science Association*, 10(1):152–164.

Kosfeld, M., Sharafi, Z., González, M. S., and Zou, N. (2025). Measuring cconomic preferences with surveys and behavioral experiments. CESifo Working Paper No. 11631.

Lepper, M. R. and Greene, D. (2015). *The hidden costs of reward: New perspectives on the psychology of human motivation*. Psychology Press.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.

Lozano, L. and Reuben, E. (2022). (Re)Measuring preferences for competition. NYUAD Working paper No. 20220078.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149):976–980.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Nikiforakis, N. and Slonim, R. (2019). Trends in experimental economics (1975–2018). *Journal of the Economic Science Association*, 5(2):143–148.

Oberlo (2024). Global internet penetration rate: Statistics and trends. Accessed: 2024-02-13.

Potters, J. and Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, 87:26–33.

Proto, E., Rustichini, A., and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127(3):1351–1390.

Proto, E., Rustichini, A., and Sofianos, A. (2022). Intelligence, errors, and cooperation in repeated interactions. *The Review of Economic Studies*, 89(5):2723–2767.

Raven, J. (1936). *The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive Mental Tests Used in Genetic Studies*. PhD thesis, University of London (King's College).

Raven, J. (2003). *Raven progressive matrices*. Springer.

Reuben, E., Sapienza, P., and Zingales, L. (2024). Overconfidence and preferences for competition. *The Journal of Finance*, 79(2):1087–1121.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Roth, A. E. (1995). Introduction to experimental economics. *The Handbook of Experimental Economics*, 1:3–109.

Shalvi, S. and De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences*, 111(15):5503–5507.

Sittenthaler, H. M. and Mohnen, A. (2020). Cash, non-cash, or mix? gender matters! the impact of monetary, non-monetary, and mixed incentives on performance. *Journal of Business Economics*, 90(8):1253–1284.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, 66(2):274–279.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72(5):923–955.

Thaler, R. H. (1991). The psychology of choice and the assumptions of economics. *Quasi Rational Economics*, 137:139–155.

UNESCO Institute for Statistics (2022). UIS.Stat Bulk Data Download Service. https://apiportal.uis.unesco.org/bdds/.

United Nations (2022). World Population Prospects 2022: Methodology of the United Nations population estimates and projections. DESA/POP/2022/TR/NO. 4.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13:75–98.

Appendix A. Further results

Table A5. Treatment differences in competitiveness

	Choosing competition		
	(1)	(2)	(3)
Incentivized	-0.082^{***}	-0.083^{***}	-0.083^{***}
	(0.019)	(0.019)	(0.019)
Performance	0.001	0.001	0.001
	(0.009)	(0.009)	(0.009)
Performance Beliefs	0.073^{***}	0.070^{***}	0.070^{***}
	(0.009)	(0.009)	(0.009)
Constant	0.567^{***}	0.568^{***}	0.568^{***}
	(0.016)	(0.016)	(0.016)
N	3000	3000	3000
Controls			
Income	No	Yes	Yes
Time	No	No	Yes

Notes: OLS regressions of participants' choice to compete. In all regressions, the dependent variable equals one for participants in the *Incentivized* treatment and zero for those in the *Hypothetical* treatment. *Performance* is the standarized number of correctly solved 0/1 tables. *Performance beliefs* are the standardized answers to the question *"Out of 100 randomly chosen participants in this study, how many do you think had fewer correct answers than you?"* . Income controls consist of 12 dummy variables, one for each income level. Time controls consist of the standardized number of seconds participants took to read the instructions and complete the task. Robust standard errors are reported in parentheses. Estimates are weighed to be nationally representative of the U.S. population in terms of age, gender, and education. ***, **, and * indicate statistical significance at 1%, 5%, and 10%.

Appendix B. Instructions

B.1. Treatment *Incentives*

Welcome

Welcome!

This study takes up to 20 minutes to complete and is designed by academics from New York University Abu Dhabi.

You will be compensated for participating in this study in the usual way. In addition, you may have bonus earnings up to **30 Dollars or more**.

Please read all instructions carefully as they describe how you can earn the bonus earnings. You will be asked questions to confirm that you have read the instructions. If you answer these questions incorrectly, you may be excluded from the study, and you may not be eligible for bonus earnings.

There are over 2,000 people participating in this study. At the end, **200 participants will be selected randomly to receive the bonus earnings**. If you are selected, your bonus payment will be sent directly to you in the form of Amazon vouchers.

Participation in this study is voluntary. You may refuse to participate or withdraw your consent at any time. Some questions will be about more personal topics such as body characteristics, religion, and political attitudes. For these topics, you may choose not to answer these questions if you prefer not to. All identifying information will be treated confidentially and identifying information will never be used. Non-identifiable information may be used in future research or shared with other researchers.

To continue, please share your consent to participate in this study.

[*I agree to participate*; *I do not wish to participate*]

———————————————**General instructions** ———————————————

Information about the bonus earnings

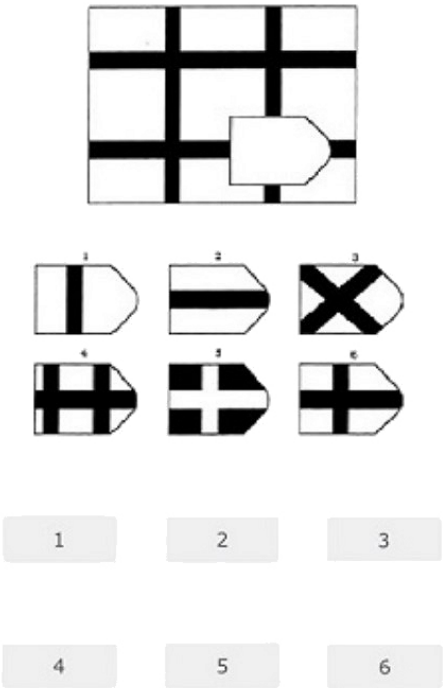
This study has six parts. If you are one of the 200 selected participants, a computer algorithm will **randomly pick one part**. Your choices in that part will be used to determine your **bonus earnings**.

The participants of this study have been carefully selected to be representative of the population of the United States aged 25 to 65.

———————————————**Cognitive skills** ———————————————

You will have to complete a test of 9 questions. Your bonus earnings in this part equal 2 Dollars per correct answer.

In the test, you will solve problems used to measure **abstract thinking**. The top part of each problem is a pattern with a piece cut out of it. Your task is to pick the piece that completes the pattern correctly. Below is an example:



In this example, the correct answer is number 6.

Get ready!

The test will start as you go to the next screen. Note: You cannot go back and forth between the problems.

page break

Problem 1

[Screenshot of Raven's matrix]

[After participants submit an answer, a new matrix appears on the screen. The sequence of matrices is the same for all participants. Participants cannot return to a previous screen. Participants have to provide an answer for all nine matrices.]

page break

The test is over.

Competitiveness

You will have **45 seconds** during which you will see tables such as the one below, filled with 1's and 0's. For each table, you will need to add up the number of 1's. For example, the correct answer to the table below is 9.

1	0	1	1
0	0	1	0
0	1	0	1
1	1	1	0

Choosing whether to compete

You can **choose how your bonus payment is calculated.** You have two options:

- **Compete**: We will compare your number of correct answers to that of a randomly chosen participant in this study. If you have more correct answers, **you win and get 2 Dollars for each correct answer**. If the other participant has more correct answers, **you lose** and you get **0 Dollars**. If there is a tie, the winner is chosen randomly.
- **Not compete**: You get **1 Dollar for each correct answer no matter how many correct answers others have**.

Understanding Question

If you answer incorrectly, you might be excluded from the study.

According to the instructions you just read, which two of the following statements are **true**? Select the two true statements.

[*With **Not Compete**, you get 1 Dollar per correct answer no matter how the other participants performed*; *With **Compete**, you get 2 Dollars per correct answer if you win the competition and zero otherwise*; *With **Compete**, you get 0 Dollars per correct answer*; *With **Not compete**, you get 0 Dollars per correct answer*][18]

——— *page break* ———

Your payment choice

Please select your payment choice.

- **Compete**: You get 2 Dollars per correct answer if you provide more correct answers than the other participant.
- **Not compete**: You get 1 Dollar for each correct answer.

Get ready!

Your 45 seconds will start as you go to the next screen.

——— *page break* ———

[18] Participants have to select the two correct answer options to proceed. They are allowed to answer the question incorrectly once. If they answer the question incorrectly a second time, they are screened out.

Table 1

[4x4 table with 0s and 1s.]

How many 1s are in this image?

To click "Next", please enter a number between 0 and 16.

[After participants submit an answer, a new table appears on the screen. The sequence of tables is the same for all participants. Participants cannot return to a previous screen. After 45 seconds, the task ends, and participants are redirected to the next screen.]

—————————————————— **Dishonesty** ——————————————————

You now have to **think of a number between 1 and 10**, and remember it. On the next screen, you will see a computer-generated number.

If the number on the screen is the same as the number you thought of, your bonus earnings from this part will be **20 Dollars**. If the numbers do not match, your bonus earnings are **10 Dollars**.

Think of a number between 1 and 10, and then click "Next."

—————————————————— *page break* ——————————————————

The computer-generated number is: [Participants see a randomly-generated number drawn from a uniform distribution with support 1 to 10]

Is the computer-generated number **the same** as the number you thought of?

- **Yes**, I thought of the same number (20 Dollars)
- **No**, I did not think of the same number (10 Dollars)

—————————————————— **Individual characteristics** ——————————————————

What is your gender?

[*Man*; *Women*; *Other*.]

What is your date of birth?

[two drop-down menus for year (*1910-2016*) and month (*January-December*)]

What is the highest level of education that you have completed?

[*No formal schooling; Primary school; Secondary school (High school); Technical/vocational training; University degree (Bachelor); Postgraduate (Masters, PhD)*]

What was your and your household's average monthly income after taxes in the last 12 months? Include income from all sources. If unsure, give us your best guesses.

[Personal income: *Less than 800 Dollars per month; 801 - 1,600 Dollars per month; 1,601 - 2,400 Dollars per month; ... ; 8,001 Dollars or more per month*; *Did not earn income in the last 12 months*; *Prefer not to answer*; Household income: *Less than 800 Dollars per month; 801 - 1,600 Dollars per month; 1,601 - 2,400 Dollars per month; ... ; 8,001 Dollars or more per month*; *No household income in the last 12 months*, *Prefer not to answer*]

B.2. Treatment *Hypothetical*

——————————————————**Welcome** ——————————————————

Welcome!

This study takes up to 20 minutes to complete and is designed by academics from a major research institution.

You will be compensated for participating in this study in the usual way. **Please read all instructions as you will be asked questions to confirm that you have read the instructions.** If you answer these questions incorrectly, you may be excluded from the study.

There are over 1,000 people participating in this study. Participation in this study is voluntary. You may refuse to participate or withdraw your consent at any time. Some questions will be about more personal topics such as body characteristics, religion, and political attitudes. For these topics, you may choose not to answer these questions if you prefer not to. All identifying information will be treated confidentially and identifying information will never be used. Non-identifiable information may be used in future research or shared with other researchers.

To continue, please share your consent to participate in this study.

[*I agree to participate*; *I do not wish to participate*]

———————————————**General instructions** ———————————————

Information about *hypothetical* bonus earnings

This study has six parts. In some parts, you will be asked to make choices to earn money as a **"bonus payment"**. Please note that while monetary amounts are presented as a currency, these amounts are **entirely hypothetical** and for the purpose of this study only. You will not be paid the bonus payment. However, even though the monetary amounts are hypothetical, please make your choices as if they are real.

The participants of this study have been carefully selected to be representative of the population of the United States aged 25 to 65.

Understanding Question

Please answer the question below. If you answer incorrectly, you might be excluded from the study.

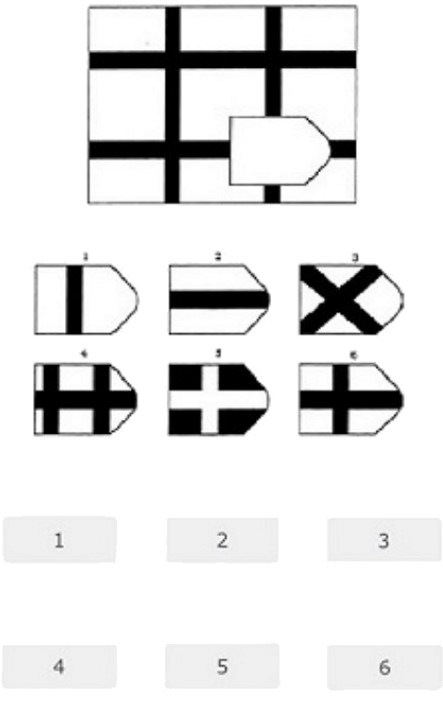
According to the instructions above, if among the 1,000 participants in this study, you are one of the 100 selected participants, which statement below is true?

[*There are six parts. In some parts you can earn bonus payments that will be paid to you at the end of the study.*; *There are six parts. In some parts you can earn bonus payments that are hypothetical and will not be paid with actual currency.*; *There are six parts. In some parts you can earn bonus payments. Some bonus payments will be paid and others are hypothetical.*][19]

Cognitive skills

You will have to complete a test of 9 questions.

In the test, you will solve problems used to measure **abstract thinking**. The top part of each problem is a pattern with a piece cut out of it. Your task is to pick the piece that completes the pattern correctly. Below is an example:



In this example, the correct answer is number 6.

Get ready!

[19]The participants are allowed to answer the question incorrectly once. If participants answer the question incorrectly a second time, they are screened out.

The test will start as you go to the next screen. Note: You cannot go back and forth between the problems.

——————————— *page break* ———————————

Problem 1

[Screenshot of Raven's matrix]

[After participants submit an answer, a new matrix appears on the screen. The sequence of matrices is the same for all participants. Participants cannot return to a previous screen. Participants have to provide an answer for all nine matrices.]

——————————— *page break* ———————————

The test is over.

——————————— **Competitiveness** ———————————

You will have **45 seconds** during which you will see tables such as the one below, filled with 1's and 0's. For each table, you will need to add up the number of 1's. For example, the correct answer to the table below is 9.

1	0	1	1
0	0	1	0
0	1	0	1
1	1	1	0

Choosing whether to compete

You can **choose how your bonus payment is calculated.** You have two options:

- **Compete**: We will compare your number of correct answers to that of a randomly chosen participant in this study. If you have more correct answers, **you win and get 2 Dollars for each correct answer**. If the other participant has more correct answers, **you lose** and you get **0 Dollars**. If there is a tie, the winner is chosen randomly.
- **Not compete**: You get **1 Dollar for each correct answer no matter how many correct answers others have**.

Understanding Question

If you answer incorrectly, you might be excluded from the study.

According to the instructions you just read, which two of the following statements are **true**? Select the two true statements.

[*With **Not Compete**, you get 1 Dollar per correct answer no matter how the other participants performed*; *With **Compete**, you get 2 Dollars per correct answer if you win the competition* *and zero otherwise*; *With* ***Compete****, you get 0 Dollars per correct answer*; *With* ***Not compete****, you get 0 Dollars per correct answer*][20]

——————————— *page break* ———————————

Your payment choice

Please select your payment choice.

- **Compete**: You get 2 Dollars per correct answer if you provide more correct answers than the other participant.
- **Not compete**: You get 1 Dollar for each correct answer.

Get ready!

Your 45 seconds will start as you go to the next screen.

——————————— *page break* ———————————

Table 1

[4x4 table with 0s and 1s.]

How many 1s are in this image?

To click "Next", please enter a number between 0 and 16.

[After participants submit an answer, a new table appears on the screen. The sequence of tables is the same for all participants. Participants cannot return to a previous screen. After 45 seconds, the task ends, and participants are redirected to the next screen.]

——————————— **Dishonesty** ———————————

You now have to **think of a number between 1 and 10**, and remember it. On the next screen, you will see a computer-generated number.

If the number on the screen is the same as the number you thought of, your bonus earnings from this part will be **20 Dollars**. If the numbers do not match, your bonus earnings are **10 Dollars**.

Remember that while bonus payments are presented as currency, these amounts are **entirely hypothetical** and for the purpose of this study only. However, even though bonus payments are hypothetical, please make your choices as if they are real.

Think of a number between 1 and 10, and then click "Next."

——————————— *page break* ———————————

[20] Participants have to select the two correct answer options to proceed. They are allowed to answer the question incorrectly once. If they answer the question incorrectly a second time, they are screened out.

The computer-generated number is: [Participants see a randomly-generated number drawn from a uniform distribution with support 1 to 10]

Is the computer-generated number **the same** as the number you thought of?

- **Yes**, I thought of the same number (20 Dollars)
- **No**, I did not think of the same number (10 Dollars)

Individual characteristics

What is your gender?

[*Man*; *Women*; *Other*.]

What is your date of birth?

[two drop-down menus for year (*1910-2016*) and month (*January-December*)]

What is the highest level of education that you have completed?

[*No formal schooling; Primary school; Secondary school (High school); Technical / vocational training; University degree (Bachelor); Postgraduate (Masters, PhD)*]

What was your [and your household's] average monthly income after taxes in the last 12 months? Include income from all sources. If unsure, give us your best guesses.

[*Less than 800 Dollars per month; 801 - 1,600 Dollars per month; 1,601 - 2,400 Dollars per month; ... ; 8,001 Dollars or more per month*; *Did not earn income in the last 12 months*; *Prefer not to answer*]