Misguided Effort[*]

Christoph Drobner[†]
A. Yeşim Orhun[‡]

First Draft: January 23, 2024

Abstract

We study how miscalibrated prior beliefs about one's own ability affect effort provision through misguided learning about returns to effort. We show that both overconfident and underconfident individuals draw misguided inferences about the returns to effort when observing initial labor market outcomes that are jointly determined by one's own ability and external luck. Importantly, we further show that this misguided learning process leads to suboptimal effort provision in the future. These results provide the first causal empirical support for a theorized effect of miscalibrated prior beliefs on economic actions that operates solely through misguided learning about the economic environment.

Keywords: Beliefs, Overconfidence, Underconfidence, Misguided Learning, Misguided Effort

JEL Codes: C91, D83, D84

[*]This experiment was pre-registered at the AEA RCT Registry (AEARCTR-0012571) and received IRB exemption status after review by the IRB Health Sciences and Behavioral Sciences at University of Michigan (HUM00231780).

[†]Technical University Munich; Email: christoph.drobner@tum.de

[‡]University of Michigan; Email: aorhun@umich.edu

1 Introduction

Imagine that you submitted a paper to a top-5 journal. You think very highly of the quality of your work but recognize that the chance of getting published also depends on the fairness of the review process. After receiving a rejection, you face a multidimensional inference problem. How should you update your beliefs about the fairness of the review process in top-5 journals and the quality of your paper? And, importantly, how do these inferences affect your future research efforts?

Psychology literature has provided ample evidence that individuals have the tendency to attribute their achievements to their own merits but their failures to external factors (see Mezulis et al., 2004, for a review).[1] Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) model such misattribution as an inference problem: miscalibrated priors about one's own ability (e.g., scholarship quality) impact learning about an external fundamental (e.g., fairness of the review process) when an individual cannot separately identify the effect of one's own ability and an external fundamental from observing one's performance (e.g., rejection). Importantly, their work examines the impact of misguided learning on economic actions and suggests that misguided learning may lead to suboptimal effort provision. Recent work in economics experimentally demonstrates that overconfidence in one's ability can trigger a misguided learning process, whereby overconfident individuals form increasingly pessimistic about external fundamentals (Goette and Kozakiewicz, 2022; Marray et al., 2020). Yet, the causal impact of misguided learning on economic actions is unexplored. Beliefs about one's potential and about the external fundamentals in one's environment, such as the returns to effort, education or investment, are fundamental inputs to decisions that may shape the course of one's life. Understanding how these (potentially misguided) beliefs influence economic decisions is key to establishing misguided learning as an economically relevant phenomenon.

In this paper, we experimentally study how miscalibrated priors about one's ability causally affect optimal actions through misguided learning about an external fundamental. Specifically, in the context of a labor market, we study how overconfidence or underconfidence about one's ability causally affects effort provision through inferences about the returns to effort. To provide evidence for the causal chain of initial belief biases impacting effort provision through misguided learning, our experiment features orthogonal interventions for both the degree of bias in priors and the direction of misguided learning conditional on that bias, being careful not to introduce a confounding path between the treatment that manipulates

[1] Examples include self-serving attribution bias in academic outcomes (Arkin and Maruyama, 1979), collective or individual performance in sport (Lau and Russell, 1980), outcomes of joint projects, for instance among couples (Ross and Sicoly, 1979).

misguided learning and the effort provision.

The experiment, illustrated by Figure 1 in Section 2, presents two periods. In the first period, participants complete a logic quiz and report their beliefs about the likelihood that they ranked in the top half of a group of four participants including themselves. We define *prior bias* as the difference between these subjective performance priors and the objective probability of ranking in the top half of a randomly drawn group of four participants. The experiment features two between-subjects variations of the quiz (Easy/ Difficult) to induce exogenous variation in subjective performance priors and the degree of prior bias. After completing the logic quiz, participants receive payments from two different evaluators: one evaluator pays participants \$2 if they score in the top half, and \$0 otherwise (performance evaluator). The other pays \$2 if a coin toss turns out to be heads, and \$0 otherwise (random evaluator). The experiment randomly assigns these roles to evaluator 1 and evaluator 2 to create exogenous variation in the direction of misguided learning that payoff feedback induces. Participants receive one of four types of payoff feedback based on their logic quiz performance and the randomly assigned evaluator roles: Both high (Evaluator 1: \$2, Evaluator 2: \$2); Mixed 1 (Evaluator 1: \$2, Evaluator 2: \$0); Mixed 2 (Evaluator 1: \$0, Evaluator 2: \$2); Both low (Evaluator 1: \$0, Evaluator 2: \$0), but do not know whether evaluator 1 or evaluator 2 is the performance evaluator. Therefore, participants receiving mixed payoff feedback in period 1 cannot separately identify the contribution of their logic quiz performance and external luck, paralleling situations that Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) study.

In the second period, participants who received mixed payoff feedback (and therefore do not know whether they scored in the top half or not) continue to work on real effort tasks. Crucially, participants are told that their payment in the second period would be determined by the evaluator 1 from the first period. If evaluator 1 is the performance evaluator, participants receive \$0.1 for each correctly solved decoding task, otherwise they receive no payment for their work in period 2. Before participants start working in period 2, they report their beliefs about the likelihood that they are paid by the performance evaluator (i.e., the likelihood that the returns to effort are positive). We define *misguidedness* as the difference between these subjective returns to effort beliefs and objective inferences Bayesian participants would have constructed if they held accurate performance priors. Finally, participants solve up to 25 decoding tasks. Their effort provision is a choice: they can choose to stop working at any time. Section 2 provides further details about the experimental design and protocol.

We predict that overconfident (underconfident) individuals attribute the high (low) payoff to the performance evaluator and the low (high) payoff to the random evaluator. Therefore, we expect overconfident participants to be positively (negatively) misguided if they are in the MIXED 1 (MIXED 2) group, and expect the opposite pattern for underconfident participants. We also predict positively (negatively) misguided individuals to over (under) exert effort in solving decoding tasks, because they are more likely to expect positive (zero) returns to effort. Section 2.2 details these predictions. In testing for the causal impact of misguided learning on effort provision, we rely on the payoff feedback treatment variation (MIXED 1 / MIXED 2) as an instrument for the variation in misguidedness. The payoff feedback treatment is designed carefully to introduce variation in the direction of misguided learning but not to introduce any other differences between these two groups that may impact effort provision. Other experiments providing feedback to create variation in posterior beliefs have either provided information about participants' ability or relative standing and/or provided different levels of financial rewards. In this context, such treatments may directly impact effort provision through mood or motivational effects. Therefore, our payoff feedback treatment reveals no information about participants' relative rank in the logic quiz (see Section 2.2 for details) and keeps the total payoffs in period 1 constant across the two groups. This design feature eliminates alternative paths between the payoff feedback treatment and outcome variables, facilitating a causal interpretation of group differences in misguidedness and effort provision across MIXED 1 and MIXED 2 groups. In addition to examining differences in the degree of misguidedness induced by payoff feedback treatment (MIXED 1/MIXED 2), we rely on its interaction with the orthogonal variation in task difficulty (EASY /DIFFICULT) to test heterogeneous theoretical predictions across feedback treatments as a function of priors.

Section 3 reports tests of our theoretical predictions regarding (1) misguided learning as a function of participants' prior bias and the experimental variation in payoff feedback, and (2) the causal impact of misguided learning on effort provision. The results confirm all of our theoretical predictions, whether we examine simple group differences or take an instrumental variables approach to deal with the possibility of unobserved confounders on the causal path from priors to effort provision. As predicted by theory, overconfident (underconfident) individuals blame the underlying fundamental (themselves) for experiencing bad labor market outcomes and consequently put less (more) effort in the next task. Individuals are on average overconfident about their abilities. Thus, we find that individuals in the MIXED 1 group (who all experience a good initial outcome) expect higher returns to effort than individuals in the MIXED 2 group (who all experience a bad initial outcome), on average. We then confirm the theorized heterogeneous difference in inferences across the MIXED 1 and MIXED 2 groups as a function of prior bias: overconfident individuals are positively misguided by 28.8 percentage points in MIXED 1 and negatively misguided by 20.5 percentage points in MIXED 2; and, underconfident individuals are negatively misguided by 14.5 percentage points in Mixed 1 and positively misguided by 19.1 percentage points in Mixed 2. Finally, we show that misguided learning has a substantial impact on effort provision. Results suggest that a 10 percentage point increase in misguidedness leads to a causal increase in the second period effort equivalent to 11% of the median number of solved decoding tasks and 8% of the median work time. The amount of misguided effort is substantial. In absolute terms, participants deviate on average by 9.3 solved decoding tasks from the optimal number of solved decoding tasks, which corresponds to almost one standard deviation from the number of solved decoding tasks.

Our results provide the first empirical support for a causal effect of prior biases on economic actions that operates through misguided learning, as theorized by Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021). In doing so, it contributes to three different strands of literature. First, it extends the burgeoning literature documenting misguided learning (Goette and Kozakiewicz, 2022; Marray et al., 2020) beyond providing additional evidence for misattribution and by employing an experimental design that addresses potential confounding relationships. Second, it complements a large body of work exploring the *direct relationships* between miscalibrated priors about one's ability and economic actions by showing evidence for the *indirect consequences* of overconfidence or underconfidence on economic behavior through misguided learning about the economic environment.[2] Third, it contributes to the broader literature examining the relationship between beliefs and actions. Barron and Gravert (2022) summarize that the existing evidence reveals a complex picture, documenting many instances in which beliefs do not affect behavior in the manner predicted by the standard model. We document a causal impact of beliefs on actions that is unconfounded by experimental design features and aligned with predictions of the standard model. Finally, from a methodological perspective, our experimental design provides a portable paradigm that shifts beliefs about one dimension, while eliminating differences in inferences about another. We hope that our work will be broadly useful for researchers studying the impact of inferences on choices in contexts that present multidimensional uncertainty.

[2]One strand of this literature emphasizes the evolutionary benefits of overconfidence (Bénabou and Tirole, 2005), such as its ability to motivate individuals to exert greater effort (Chen and Schildberg-Hörisch, 2019) or to persuade others (Schwardmann and Van der Weele, 2019; Schwardmann et al., 2022; Solda et al., 2020). Conversely, another strand in this literature highlights the detrimental consequences of overconfidence (Malmendier and Taylor, 2015), including excessive risk-taking in financial markets (Barber and Odean, 2001), sub-optimal managerial decisions (Malmendier and Tate, 2005), and selection into competition (Camerer and Lovallo, 1999; Niederle and Vesterlund, 2007)

2 Experimental Design and Theoretical Predictions

Individuals are known to hold biased beliefs about their ability (see Moore and Healy, 2008, for a review). Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) theoretically show that when individuals cannot separately identify how much their ability versus external fundamentals (e.g., returns to effort) contribute to their productivity, they will learn misguidedly about the external fundamental and make suboptimal choices.[3] Misguided learning is consistent with Bayesian updating, and particularly strong when beliefs about ability do not converge to the truth (Heidhues et al., 2018). Below, we first detail how the experiment is designed to parallel this context and discuss theoretical predictions regarding misguided learning and effort provision. Then, we explain how the experimental design allows us to provide causal evidence for these predictions. We conclude this section with procedural details. All experimental instructions are provided in Appendix C.

2.1 Overview

The experiment, illustrated by Figure 1, consists of two periods. In both periods, participants work to solve tasks. At the end of the experiment, one period is chosen to determine payments. In both periods, expected payoffs are increasing in the number of correctly solved tasks.

Period 1 In period 1, participants perform a logic quiz with 12 puzzles from Civelli et al. (2018) that are similar to the Raven Progressive Matrix test (a commonly used test to measure fluid intelligence), learn their score and report their beliefs $\gamma \in (0, 1)$ about the likelihood of having performed in the top half in a group of 4 participants (including themselves) in the logic quiz.[4] We refer to these beliefs γ as *performance priors.* Figure 1 denotes measurements collected from participants within ellipses on the left.

Then, participants receive feedback that is jointly determined by their relative performance on the logic quiz and an external fundamental that determines returns to effort. The feedback comes in the form of two payoffs (s_1, s_2) for participants' performance in period 1. One payoff is provided by an evaluator who pays participants based on their relative performance in the logic quiz: if the participant's performance ranks in the top half of the

[3] Van den Steen (2004) also studies a similar setting, but overconfidence is generated by task choice: people choose tasks they overestimate their chance of success in, and when they fail, they attribute it externally, thus optimism lives on. Without task choice, there is no overconfidence.

[4] We incentivize these performance prior reports with the binarized scoring rule without providing detailed information about the incentives (as in Danz et al., 2022). Participants know that they have the chance to win a $1 bonus, and the chance of winning this bonus increases with the accuracy of their performance prior reports.

group of 4 participants, the evaluator pays H, and otherwise pays L. We call this evaluator the performance evaluator. The other payoff is provided by an evaluator who tosses a coin to pay H if the coin toss is heads and L otherwise. We call this evaluator the random evaluator. We set H to \$2 and L to \$0. Participants receive both payoffs as payment, and know how much evaluator 1 and evaluator 2 paid them, but do not know the roles of either evaluator. This information asymmetry is emphasized by the dashed lines around the box containing the treatment assignment in Figure 1.

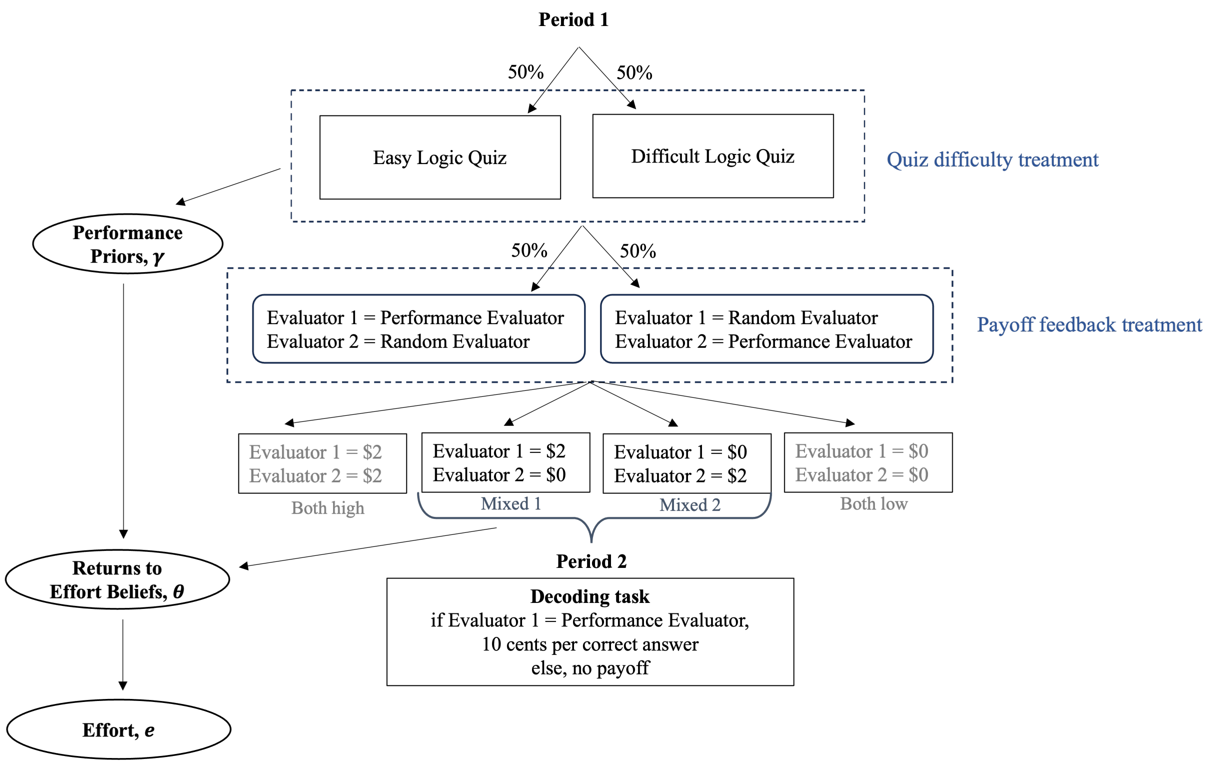


Figure 1: Experimental Design

Treatments The experiment features two between-subjects binary treatments that are assigned orthogonally and with equal probability. The first treatment assigns participants to Easy and Difficult versions of the logic quiz. While questions 1-7 are the same among both conditions, questions 8-12 vary in difficulty levels between the conditions. The purpose of this manipulation is to provide exogenous variation in participants' beliefs about their relative performance in the logic quiz.

The second treatment randomizes the evaluator roles assigned evaluator 1 and evaluator 2. Denoting s_i as the signal from evaluator $i \in 1, 2$, participants receive one of four types of feedback: Both High ($s_1 = H, s_2 = H$), Mixed 1 ($s_1 = H, s_2 = L$), Mixed 2 ($s_1 = L, s_2 =$

H), Both Low ($s_1 = L, s_2 = L$), based on their logic quiz performance and the randomly assigned evaluator roles. Participants who receive only high payoffs or low payoffs from both evaluators learn whether their performance in the logic quiz ranks among the top half and therefore are redirected to a different survey. The analysis sample consists of participants in Mixed 1 and Mixed 2 groups, for whom the feedback provides no information about one's relative ability. We refer to the exogenous experimental variation in signals between Mixed 1 and Mixed 2 as the payoff feedback treatment.

Period 2 In period 2, participants are presented with 25 decoding tasks. Each task involves translating a 5 digit number into text based on the decoding key. Participants are informed that they can decide whether they want to continue or stop working after each decoding task. This stopping option is introduced to increase the responsiveness of effort provision to monetary incentives by making the opportunity cost of working on the decoding task salient.[5]

Participants are also informed that they are paid for their work in period 2 by the same evaluator 1 from period 1. Thus, the external fundamental is constant across periods 1 and 2. If evaluator 1 is the performance evaluator, the returns to effort are positive: participants receive a piece-rate ω (set to \$0.1) for each correctly solved decoding task.[6] If evaluator 1 is the random evaluator, they receive no payoff for their work in period 2. Before participants start working on the decoding tasks, we elicit their beliefs $\theta \overset{\text{def}}{=\!=} pr(P) \in (0,1)$ about the likelihood that evaluator 1 is the performance evaluator (P).[7] We refer to these beliefs θ as *returns to effort beliefs* regarding the period 2 task. Finally, we observe the number of decoding tasks participants choose to work on, which we refer to as effort provision, e.

2.2 Theoretical Predictions

We define prior bias ($\Delta\gamma$) as the difference between subjective performance priors (γ) and the objective probability of ranking in the top two among a group of four participants (γ^*).

Definition 1 prior bias $\Delta\gamma \overset{\text{def}}{=\!=} \gamma - \gamma^*$

[5]Several other studies increase the salience of the opportunity cost of working on a real effort task in order to increase responsiveness of effort provision to monetary incentives (see, e.g., Chen and Schildberg-Hörisch, 2019; DellaVigna et al., 2022; Erkal et al., 2018; Goerg et al., 2019).

[6]We deliberately choose a piece-rate instead of a tournament-based performance payoff in period 2. This design feature ensures that participants' effort provision in period 2 depends on their beliefs about the returns to effort, but eliminates confounds arising from strategic considerations regarding other's beliefs about the returns to effort and their effort provisions.

[7]We elicit these beliefs after providing information about payments in period 2, because we recognize that these instructions might influence beliefs about evaluator 1's type above and beyond what is implied by their performance priors.

Note that prior bias can take on any value between -1 and 1. When prior bias is positive (negative), the participant is overconfident (underconfident) about their performance.

Now, we turn to the impact of the payoff feedback treatment. Note that, a priori, there is a 50% chance that evaluator 1 is the performance evaluator. However, given priors γ, participants make Bayesian inferences about the probability that evaluator 1 is the performance evaluator from the payoff feedback to construct returns to effort beliefs θ. The participants in MIXED 1 observe the payoff feedback ($s_1 = H$, $s_2 = L$). A Bayesian subject constructs the following beliefs about the likelihood that evaluator 1 is the performance evaluator (P) when $s_1 = H$:

$$\begin{aligned} pr(P|s_1 = H) &= \frac{pr(P)Pr(s_1 = H|P)}{pr(P)Pr(s_1 = H|P) + (1 - pr(P))(1 - Pr(s_1 = H|P))} \\ &= \frac{0.5\gamma}{0.5\gamma + .5(1-\gamma)} = \gamma \end{aligned} \tag{1}$$

Participants in MIXED 2 observe the payoff feedback ($s_1 = L$, $s_2 = H$). Bayesian returns to effort beliefs when $s_1 = L$ are:

$$\begin{aligned} pr(P|s_1 = L) &= \frac{pr(P)Pr(s_1 = L|P)}{pr(P)Pr(s_1 = L|P) + (1 - pr(P))(1 - Pr(s_1 = L|P))} \\ &= \frac{0.5(1-\gamma)}{0.5(1-\gamma) + .5\gamma} = 1 - \gamma \end{aligned} \tag{2}$$

Therefore, as in Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021), the information extracted from feedback regarding returns to effort depends on participants' performance priors. In the MIXED 1 treatment, θ is predicted to equal γ and in the MIXED 2 treatment, θ is predicted to equal $1 - \gamma$. This inference is misguided to the extent that performance priors γ depart from the objective probability of ranking in the top half, γ^*. Denote objective returns to effort beliefs with θ^*. Simply applying Bayes' rule to the objective probability of that individual ranking in the top half, $\theta^* = \gamma^*$ if $s_1 = H$ and as $\theta^* = 1 - \gamma^*$ if $s_1 = L$. We define misguidedness ($\Delta\theta$) as the difference between the participant's returns to effort belief and the objective returns to effort belief.

Definition 2 MISGUIDEDNESS $\Delta\theta \stackrel{\text{def}}{=} \theta - \theta^*$

Note that misguidedness ranges from -1 to 1. When individuals are positively (negatively) misguided, they are more optimistic (pessimistic) about returns to effort than they should be. For Bayesian individuals, misguidedness is equal to $\gamma - \gamma^* = \Delta\gamma$ if $s_1 = H$ and it is equal to $\gamma^* - \gamma = -\Delta\gamma$ if $s_1 = L$.[8] Thus, we obtain our main hypothesis pertaining to misguided learning in the context of our experimental design:

Hypothesis 1 *If individuals are initially overconfident ($\Delta\gamma > 0$), they will be positively misguided in* MIXED 1 *and negatively misguided in* MIXED 2*. If individuals are initially underconfident ($\Delta\gamma < 0$), they will be negatively misguided in* MIXED 1 *and positively misguided in* MIXED 2*. More generally, the differences in misguidedness between* MIXED 1 *and* MIXED 2 *increase monotonically in prior bias.*

Finally, we consider the effect of misguided learning on effort provision in period 2. The expected utility from exerting effort e depends on expected returns to effort provision, $\theta\omega e$, and the cost of effort, $c(e)$:

$$u(e, \theta) = \theta\omega e - c(e) \tag{3}$$

Given the impact of θ on expected utility maximizing effort, effort provision is suboptimal to the extent that θ departs from the objective probability θ^*. We define misguided effort (Δe) as the difference between the participants effort and the level of effort participants would have exerted, if their inferences regarding returns to effort were not misguided (i.e., their beliefs were θ^*).

Definition 3 MISGUIDED EFFORT $\Delta e \stackrel{\text{def}}{=\!=} argmax_e(u(e, \theta))$ - $argmax_e(u(e, \theta^*))$

Assuming that $c(e)$ is convex in e, we can write our main hypothesis pertaining to effort provision as:

Hypothesis 2 *The expected utility-maximizing effort is monotonically increasing in θ and consequently monotonically increasing in misguidedness $\Delta\theta$. The effort provision of positively (negatively) misguided individuals is higher (lower) than the effort they would have provided if they were not misguided.*

2.3 Causal Inference

In the context of this experiment, theory predicts (1) prior bias to cause misguided learning about returns to effort, and (2) misguided learning about returns to effort to cause distortions in effort provision. This causal chain from prior bias to effort provision through misguided learning is depicted in Figure 2. We discuss how the experimental design allows us to provide causal empirical evidence for these predictions.

[8] In Appendix B, we show that participants closely follow Bayesian updating and discuss any empirical departures.

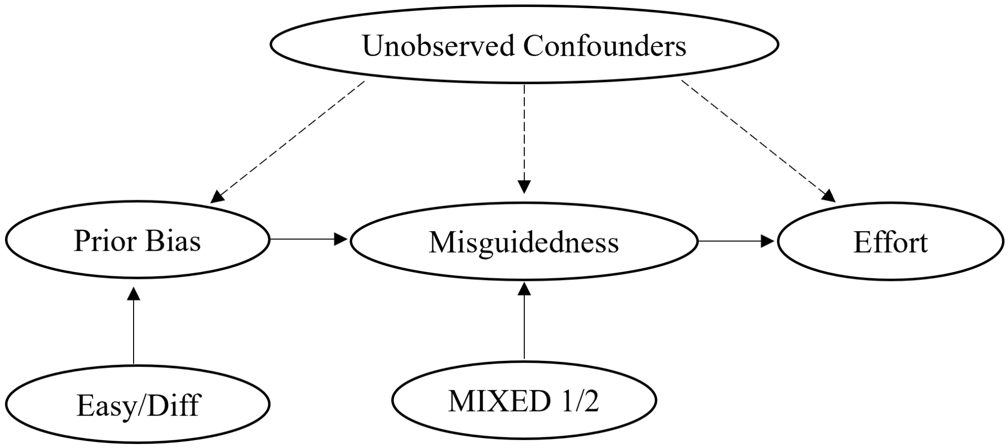


Figure 2: Representation of the Causal Mediation Chain and Treatments

First, note that unobserved confounders may cause spurious correlations between the variables on the causal chain, which creates a challenge for causal inference when these variables are measured rather than manipulated. For example, the potential to performing well on the experimental tasks (due to one's intelligence, concentration, or motivation to perform, etc.) is an unobserved individual factor that creates a spurious correlation between logic quiz scores (and therefore, prior bias) and decoding task performance. This factor can also correlate with belief updating (i.e., how participants construct inferences given their priors and the payoff feedback), also leading to a confounding path between the degree of misguidedness and performance (on the logic quiz and the decoding task). There may also be other unobserved confounders.

The experimental paradigm allows us to identify the causal path even in the presence of unobservable confounders, as randomized treatments can be used as instruments for the variables on the causal chain. In particular, the Easy/Difficult quiz difficulty treatment provides an instrument for the prior bias, as we expect participants to form more optimistic beliefs about their relative performance in the Easy condition compared to the Difficult condition (Moore and Healy, 2008). The Mixed 1/Mixed 2 payoff feedback treatment provides an instrument for misguidedness, as we expect $\theta = \gamma$ in Mixed 1 and $\theta = 1 - \gamma$ in Mixed 2. In line with the validity requirements for instruments, the representation in Figure 2 specifies that the quiz difficulty treatment has a direct impact on prior bias, but no direct effect on misguidedness. Similarly, the payoff feedback treatment has a direct impact on misguidedness, but no direct effect on effort. The lack of a direct effect of the instrument on the outcome variable of interest is known as the excludability assumption.

Feedback treatments often create variation in earnings or provide information about relative performance or ability. Such treatments would likely violate the excludability assumption since variations in earnings or ego-relevant information can directly impact effort provision through creating differential mood or motivational effects. A contribution of this paper is to design a feedback treatment that does not create such variations. Note that both Mixed 1 and Mixed 2 groups receive \$2 from evaluator 1 and 2 combined. To see that no information about performance is transmitted by the payoff feedback treatment to participants in the Mixed 1 or Mixed 2 groups, denote γ_0 as the participant's prior performance beliefs before observing payoff signals. The probability of evaluator 1 being the random evaluator is 0.5. By definition, $Pr(s_i = H|H) = 1$ when the evaluator i is a performance evaluator, and $Pr(s_i = H|H) = 0.5$ when the evaluator i is a random evaluator. Thus, for any evaluator i, $Pr(s_i = H|H) = 0.75$ and $Pr(s_i = L|H) = 0.25$, resulting in $Pr(s_i = H, s_{-i} = L|H) = 0.5$. Consequently,

$$\begin{aligned}\gamma_1(s_i = H, s_{-i} = L) &= \frac{Pr(s_i = H, s_{-i} = L|H)\gamma_0}{Pr(s_i = H, s_{-i} = L|H)\gamma_0 + (1 - Pr(s_i = H, s_{-i} = L|H))(1 - \gamma_0)} \\ &= \frac{0.5\gamma_0}{0.5\gamma_0 + 0.5(1 - \gamma_0)} = \gamma_0\end{aligned} \tag{4}$$

This design feature is crucial to defend the plausibility of the excludability assumption, as it allows us to rule out potential confounding effects that learning ego-relevant information or payment level variations may have on future performance. It also brings the experimental design in parallel with the setting considered by Heidhues et al. (2018) where individuals do not update their beliefs about their ability.

2.4 Procedures

We programmed the experiment with *Qualtrics*. We recruited 2,011 participants from the US on *Prolific*, half of them women. A total of 1,004 participants received mixed payoff feedback in period 1 and consequently completed period 2. Our analyses in Section 3 focus on these participants. Participants received a completion fee of \$2 and the bonus payment from one randomly chosen period of the experiment. The average earnings were \$3.7.

2.5 Block Randomization and Balance

Exactly half of the 1,004 participants in the analysis sample are in the Mixed 1 group (254 of them had solved the Easy quiz in Period 1), and the other half are in the Mixed 2 group (251 of them had solved the Easy quiz in Period 1). Randomization of the logic quiz difficulty treatment was stratified across gender. The randomization of the payoff feedback treatment was stratified across quiz difficulty, deciles of performance beliefs, gender, outcome of the coin toss and whether or not the participant's performance is in the top half of the score distribution. The data is therefore balanced across treatments with respect to these variables (see Online Appendix, Table A).

As we discuss above, our experimental design allows us to provide causal evidence for misguided learning and misguided effort even in the presence of unobserved confounders, such as one's potential of performing well on experiment tasks. We also calculate the number of correct answers on the common questions in the logic quiz (i.e., the first seven questions) as a proxy for performance potential and find that the treatment groups are balanced with respect to this common questions score (4.95 in Mixed 1, 4.85 in Mixed 1, $p = 0.483$; 4.86 in Easy, 4.94 in Difficult, $p = 0.345$). This result suggests that the randomization produced comparable treatment groups in terms of individuals' performance potential distribution.

3 Results

We begin with presenting definitions and summary statistics of the main variables used in our analyses in Section 3.1. The empirical findings are organized into three main subsections that follow. In Section 3.2, we examine the impact of the quiz difficulty treatment (Easy/Difficult) on performance priors and prior bias. In Section 3.3, we analyze how participants' prior bias and the payoff feedback treatment (Mixed 1/Mixed 2) leads to misguided learning about the returns to effort, testing hypothesis 1. Recall that the experimental variation in prior bias generated by the quiz difficulty treatment allows us to identify the causal heterogeneous treatment effect of the payoff feedback treatment on the degree of misguided learning. Finally, in Section 3.4, we analyze the consequences of this misguided learning for participants' efforts on the decoding tasks in period 2, testing hypothesis 2. Recall that the experimentally manipulated payoff feedback treatment (Mixed 1/Mixed 2) exogenously moves inferences regarding returns to effort independent of any unobserved confounders such that we can test for the causal impact of misguided learning on participants' subsequent effort provision.

3.1 Overview

The summary statistics of the main variables informing our analyses are presented in Table 1. On average, participants correctly solve 7.9 logic quiz questions (logic score) in total, and 4.9 questions among the seven questions that are common across the Easy and Difficult logic quiz versions (common questions score). Performance priors reflect aggregate overconfidence. Participants expect to rank among the top two in a group of randomly selected four participants 59.1% of the time and have an average prior bias of 8.6% (Wilcoxon signed-rank tests, $p < 0.001$). Recall that we define prior bias as the difference between a participant's performance prior and the objective probability of that participant ranking in the top half. We calculate this objective probability by repeatedly drawing groups of three from the sample of individuals who completed the logic quiz and calculating the fraction of times the participant's logic quiz score exceeds at least two of the three comparison individuals. It is important to note that owing to the block random assignment mechanism, performance priors and prior bias are indistinguishable across the Mixed 1 and Mixed 2 groups (performance priors are 59.37 vs 58.80, respectively, contrast $p = 0.696$; prior bias is 8.56 vs 8.59, contrast $p = 0.894$). However, there is a great deal of variation in performance priors: 26.7% of the individuals believe that the chances of scoring in the top half are lower than 50%, and the absolute value of the bias is on average 24.1%. We will return to this variation in the next section.

Table 1: Summary Statistics of the Sample

Variable Name	Mean	S.D.	Min	q25	Median	q75	Max
Logic Score	7.9	2.5	0	6	8	10	12
Common Questions Score	4.9	1.6	0	4	5	6	7
Performance Priors	59.1	25.9	0	40	60	80	100
Prior Bias	8.6	29.4	-78.6	-9.9	8.2	26.8	97.1
Abs. Prior Bias	24.1	18.9	0	9.7	18.7	36.8	97.1
Returns to Effort Beliefs	54.0	27.0	0	40	50	75	100
Misguidedness	3.7	37.6	-100	-22.2	1.8	32.5	99.9
Abs. Misguidedness	30.4	22.4	0	11.8	26.7	46.2	100
Decoding Tasks Solved	12.8	10.0	1	3	10	25	25
Decoding Work Time (seconds)	216	180	1	50	174	347	968
Misguided Tasks Solved	0.4	10.4	-16.5	-8.0	-1.4	9.2	17.7
Misguided Work Time (seconds)	5	183	-272	-140	-32	131	804
Abs. Misguided Tasks Solved	9.3	4.7	0	6.3	9.0	13.1	17.7
Abs. Misguided Work Time (seconds)	153	100	1	86	136	202	804

Looking at inferences regarding returns to effort, we see that participants expect the evaluator in Period 2 to be the performance evaluator 54% of the time, and are misguided by 3.7% on average. Recall that we define misguidedness as the difference between the participant's returns to effort belief and the objective returns to effort belief. We calculate the objective returns to belief by applying Bayes' rule to the objective probability of that individual ranking in the top half. Thus, individuals are expected to be negatively (positively) misguided on average if their prior bias is negative (positive). Looking at the absolute value of misguidedness, which averages at 30.4%, we see that the deviations are substantial.

The last set of variables in Table 1 pertain to the effort participants put forth on the decoding task in Period 2. We focus on the number of tasks solved and the amount of time spent working as measures for effort put forth by the participants. On average, participants solve 12.8 tasks out of 25 available tasks and spend 216 seconds working. However, the distribution of effort has a large variance, as we will discuss further. We calculate misguided effort (misguided tasks solved and misguided work time) as the difference between the effort participants put forth and the effort they would put forth if they held accurate performance priors. The counterfactual is based on the causal estimates of the impact beliefs have on effort provision in Section 3.4) and we will discuss it in further detail below. Here, we want to highlight that absolute levels of misguided effort are quite substantial: on average, effort provision deviates from the counterfactual levels by 9.3 tasks and 153 seconds.

3.2 Performance Priors and Prior Bias

We first confirm that the Easy/Difficult successfully manipulates performance priors and prior bias. Panel A of Figure 3 plots the distribution of performance priors of Easy quiz participants with red bars and the distribution of performance priors of Difficult quiz participants with blue bars. Panel B does the same for prior bias. The distributions of performance priors and prior bias in the Easy quiz condition are shifted to the right of those in the Difficult quiz condition (Kolmogorov–Smirnov test, both $p's < 0.001$).

On average, participants in the Easy quiz condition hold performance priors of 66.7%, while participants in the Difficult quiz condition hold performance priors of 51.4%, leading to a difference of 15.3 percentage points (Wilcoxon rank-sum test, $p < 0.001$). Participants in the Easy quiz condition are substantially overconfident, with an average prior bias of 15.4% (Wilcoxon signed-rank test, $p < 0.001$). Participants in the Difficult quiz condition are not significantly overconfident, with an average prior bias of 1.7% (Wilcoxon signed-rank test, $p = 0.274$). The 13.7 percentage points difference in prior bias between Easy quiz and Difficult quiz conditions is statistically significant (Wilcoxon rank-sum test, $p < 0.001$).[9] Performance priors and prior bias are balanced between Mixed 1 and Mixed 2 groups

[9] Participants solve more questions correctly in the easy logic quiz than in the difficulty one (9.3 versus 6.6, $p < 0.001$). A participant's logic quiz score is a strong predictor for their performance priors: on average, a 1 point increase in logic quiz scores is associated with a 6.7 percentage points increase in performance priors ($p < 0.001$).

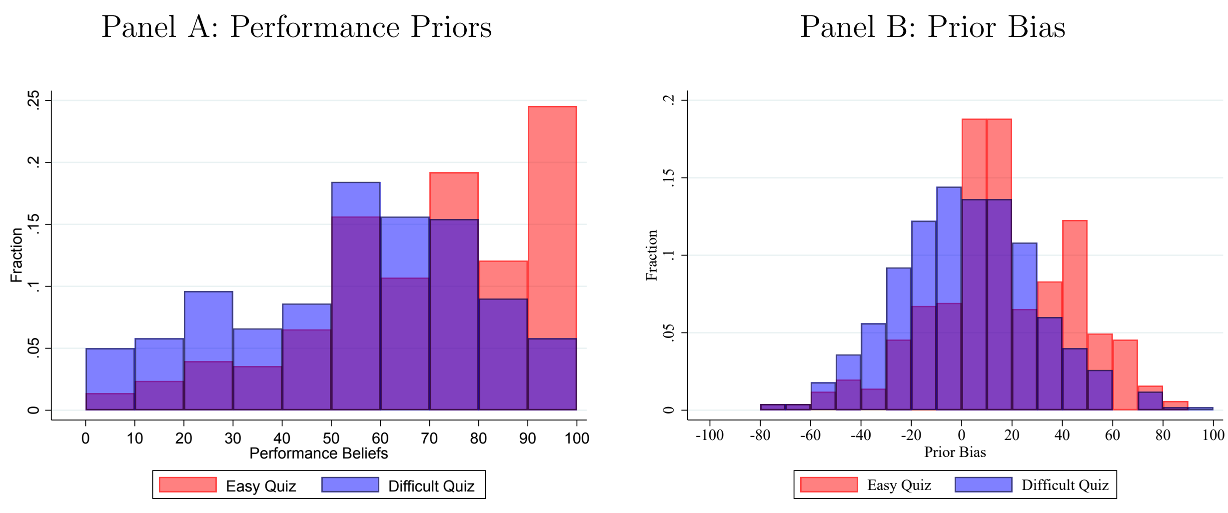


Figure 3: Manipulation check of the quiz difficulty treatment

($p's > 0.695$). Next, we examine the downstream consequences of performance priors and prior bias on inferences about the returns to effort.

3.3 Returns to Effort Beliefs and Misguided Learning

Hypothesis 1 predicts that overconfident individuals are positively misguided in Mixed 1 and negatively misguided in Mixed 2 whereas underconfident individuals are negatively misguided in Mixed 1, and positively misguided in Mixed 2. Because participants are overconfident overall, we first provide aggregate results. Figure 4 contrasts the means and 95% confidence intervals of returns to effort beliefs (Panel A) and the degree of misguided learning (Panel B) across Mixed 1 and Mixed 2 groups.

Panel A in Figure 4 shows that participants in the Mixed 1 group believe that effort pays off with a probability of 64.0%, while participants in the Mixed 2 group believe that effort pays of with a probability of 43.9%, leading to a difference of 20.1 percentage points (Wilcoxon rank-sum test, $p < 0.001$). Panel B in Figure 4 shows that participants are positively misguided by 13.2 percentage points in the Mixed 1 group (Wilcoxon signed-rank test, $p < 0.001$) and negatively misguided by 5.9 percentage points in the Mixed 2 group (Wilcoxon signed-rank test, $p < 0.001$), leading to a difference of 19.1 percentage points (Wilcoxon rank-sum test, $p < 0.001$).

Next, we plot average misguidedness levels across Mixed 1/Mixed 2 conditions for overconfident ($\Delta\gamma > 0$) individuals (Panel A) and for underconfident ($\Delta\gamma \leq 0$) individuals (Panel B) in Figure 5. Panel A in Figure 5 shows that overconfident individuals are positively misguided about the returns to effort by 28.8 percentage points in Mixed 1 (Wilcoxon

Panel A: Returns to Effort Beliefs

Panel B: Misguided Learning

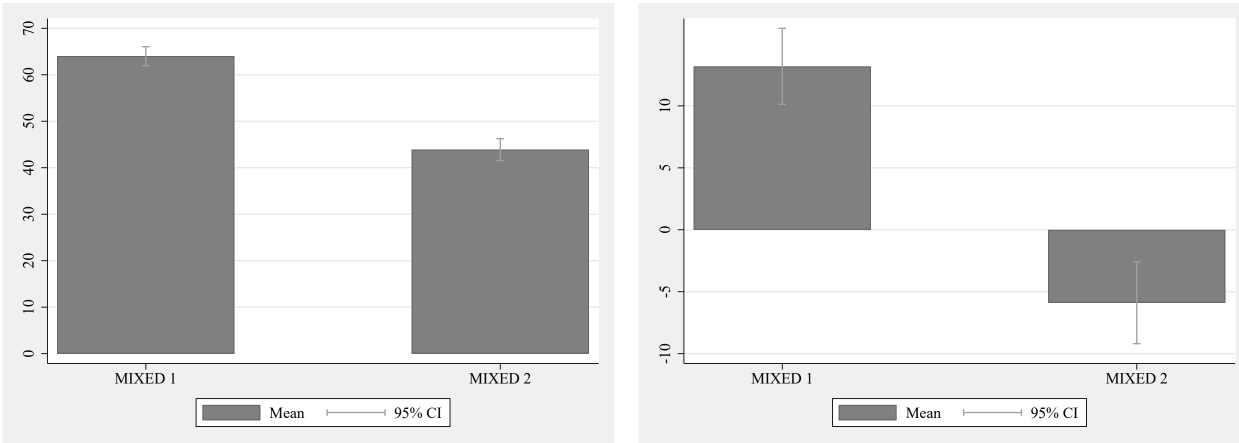


Figure 4: Aggregate Returns to Effort Beliefs and Misguided Learning across Mixed 1 and Mixed 2 groups

signed-rank test, $p < 0.001$) while they are negatively misguided about the returns to effort by 20.5 percentage points in Mixed 2 (Wilcoxon signed-rank test, $p < 0.001$). The theorized opposite pattern emerges in Panel B of Figure 5 for underconfident participants. Specifically, underconfident participants are negatively misguided about the returns to effort by 14.5 percentage points in Mixed 1 (Wilcoxon signed-rank test, $p < 0.001$) while they are positively misguided about the returns to effort by 19.1 percentage points in Mixed 2 (Wilcoxon signed-rank test, $p < 0.001$).

Panel A: Overconfident Individuals

Panel B: Underconfident Individuals

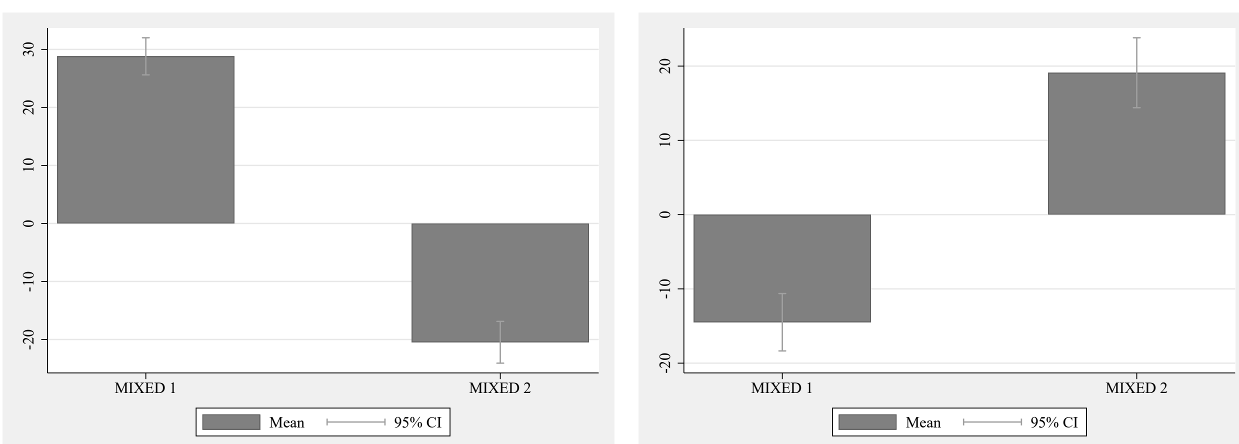


Figure 5: Misguided Learning across Mixed 1 and Mixed 2 groups, Split by the Direction of Prior Bias

Hypothesis 1 more generally predicts that the differences in the degree of misguided learning between Mixed 1 and Mixed 2 increase monotonically in prior bias. To directly test for the hypothesized heterogeneous treatment effects, we first estimate an OLS specification that regresses our measure of misguided learning on an indicator for Mixed 1, the degree of prior bias and the interaction of the two. The coefficients are reported in Column 1 of Table 2 and confirm that the difference in misguided learning between Mixed 1 and Mixed 2 groups is increasing in prior bias.

Table 2: Testing the Hypothesized Heterogenous Treatment Effect of Mixed 1/Mixed 2 Manipulation

Dependent Variable:	Misguided Learning		Returns to Effort Beliefs	
	(1)	(2)	(3)	(4)
Mixed 1	19.088***	19.091***	20.061***	20.026***
	(1.704)	(1.896)	(1.433)	(1.432)
Prior Bias	-0.759***	-0.228		
	(0.043)	(0.226)		
Mixed 1*Prior Bias	1.655***	0.771***		
		(0.053)	(0.279)	
Prior			-0.312***	-0.292***
			(0.050)	(0.155)
Mixed 1*Prior			0.808***	0.892***
			(0.064)	(0.190)
Constant	-5.876***	-3.929		
	(1.352)	(2.587)		
Observations	1,004	1,004	1,004	1,004
Instrumental Variables	No	Yes	No	Yes

Notes:
(i) Results reported in columns 1 and 3 are derived from OLS regressions. Results reported in columns 2 and 4 are derived from 2SLS regressions where the potentially endogenous variables that are functions of performance priors are instrumented by a dummy indicating quiz difficulty and the interaction between dummies for the quiz difficulty and the payoff feedback treatments. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

However, note that performance priors and consequently prior bias are measured variables. A causal interpretation of the heterogeneity result is limited by potential unobservable confounders that influence both performance priors and inferences about the returns to effort. Therefore, we rely on the exogenous variation in performance priors created by the experimental design to provide a valid test of hypothesis 1.

In particular, we estimate a two-stage least squares regression where quiz difficulty treatment and the interaction between the quiz difficulty and the payoff feedback treatments are used as instruments for the potentially endogenous variables that are functions of performance priors. The result of this 2SLS specification for misguided learning is reported in Column 2 of Table 2. We find that a 10 percentage point increase in prior bias on average leads to a 7.7 percentage points increase in the difference of misguided learning across the Mixed 1 and Mixed 2 groups ($p < 0.001$). This result provides strong causal evidence in support of hypothesis 1.

Although hypothesis 1 focuses on misguided learning, based on the preliminaries presented in Section 2.2 it is easy to confirm that we would predict that the difference in returns to effort beliefs across the Mixed 1 and Mixed 2 groups to increase in performance priors. Columns 3 and 4 of Table 2 report the corresponding OLS and 2SLS regression results to test this conjecture. Again, we find evidence in support of the theoretical predictions: a 10 percentage point increase in performance priors on average leads to a 8.1 percentage points increase in the difference of returns to effort beliefs across Mixed 1 and Mixed 2 groups ($p < 0.001$).

Overall, the results show that miscalibrated performance priors lead to misguided learning about the underlying fundamental: participants in the Mixed 1 group misguidedly infer higher returns to effort than participants in the Mixed 2 group. Furthermore, the treatment differences in returns to effort beliefs and misguided learning are moderated by prior bias. Next, we examine the consequences of returns to effort beliefs and misguided learning on effort provision in period 2.

3.4 Misguided Effort

Hypothesis 2 predicts that returns to effort beliefs and the degree of misguided learning have a positive and causal effect on effort provision. Recall from Section 3.3 that participants in the Mixed 1 group hold ceteris paribus more positive beliefs about the returns to effort than participants in the Mixed 2 group. Therefore, a positive difference in effort provision between Mixed 1 and Mixed 2 groups would provide preliminary evidence that returns to effort beliefs and misguided learning causally affect effort provision.

Panel A in Figure 6 shows the distributions of solved decoding tasks separately for Mixed 1 and Mixed 2 groups. We see that the distribution is bi-modal: most people either quit after working one or a few decoding tasks or work on all 25 tasks. Participants in the Mixed 1 group were disproportionately more likely to work on all 25 decoding tasks in period 2 (Fisher's exact test, $p = 0.011$), and participants in the Mixed 2 group were disproportionately more likely to work on the minimum of 1 decoding task (Fisher's exact

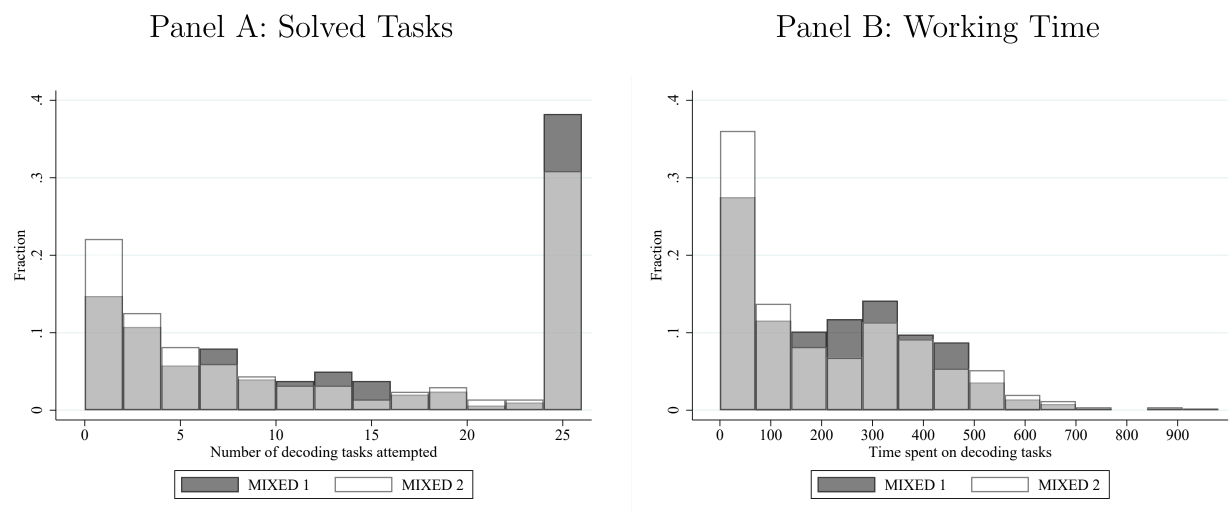


Figure 6: Effort Provision in the Decoding Task across Mixed 1 and Mixed 2 groups

test, $p = 0.003$).[10] On average, participants in the Mixed 1 group solved 13.5 decoding tasks, while participants in the Mixed 2 group solved 11.5 decoding tasks. This difference is statistically significant (Wilcoxon rank-sum test, $p < 0.001$). The difference in effort provision between Mixed 1 and Mixed 2 groups is also visible in Panel B of Figure 6, which shows the distributions of the time spent on the decoding task. On average, participants in the Mixed 1 group spent 226 seconds working on the decoding task while participants in the Mixed 2 group spent 205 seconds working on the decoding task. This difference is statistically significant (Wilcoxon rank-sum test, $p < 0.008$). Overall, the positive difference in effort provision between Mixed 1 and Mixed 2 groups provides initial evidence in support of hypothesis 2.

The differences in effort provision across participants in the Mixed 1 and Mixed 2 groups translate to substantial payoff consequences. The realized period 2 payoffs are on average 88.5 cents in the Mixed 1 group while participants in the Mixed 2 group earn 53 cents (Wilcoxon rank-sum test, $p < 0.001$). This difference in payoffs becomes starker when we compare payoffs from participants who were paid by the performance evaluator in period 2. Among them, participants in the Mixed 1 group earn on average 1.56 dollars while participants in the Mixed 2 group earn on average 1.10 dollars (Wilcoxon rank-sum test, $p < 0.001$). Of course, the higher earnings in the Mixed 1 group presumably come with higher effort costs. Next, we establish the extent to which differences in effort provision are driven by the degree of misguided learning about the returns to effort, and quantify

[10] The minimum number of solved decoding tasks was 1 because we forced our participants to work on at least one decoding task. Participants then decided after each decoding task whether they wanted to continue or stop working.

deviations from the optimal level of effort provision as a consequence of initial prior bias.

Table 3: The Causal Impact of Misguided Learning and Returns to Effort Beliefs on Effort Provision

Dependent Variable:	Decoding Tasks Solved		Decoding Time	
	(1)	(2)	(3)	(4)
Returns to Effort Beliefs	0.101***		1.362**	
	(0.029)		(0.529)	
Misguided Learning		0.108***		1.402**
		(0.033)		(0.589)
Constant	7.326***	12.505***	142.275	210.656
	(1.598)	(0.345)	(29.346)	(6.260)
Observations	1,004	1,004	1,004	1,004
Instrumental Variables	Yes	Yes	Yes	Yes

Notes:
(i) Results reported are derived from 2SLS regressions where returns to effort beliefs and the degree of misguided learning are instrumented by a full-interaction between dummies indicating quiz difficulty and payoff feedback treatments. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**\,p < 0.05$, $***p < 0.01$.

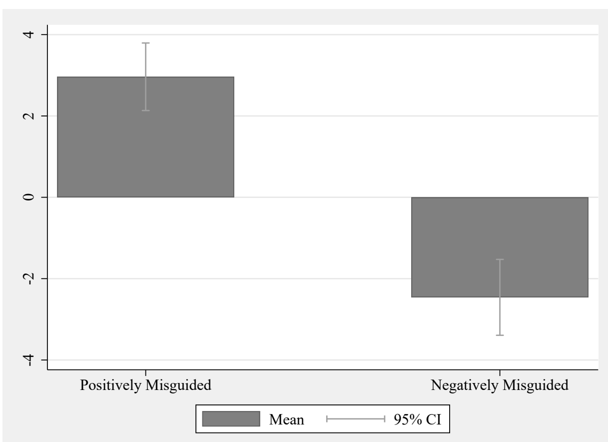
To provide causal evidence for the hypothesized positive impact of returns to effort beliefs and misguided learning on effort provision, Table 3 reports results from four two-stage least squares regressions. First, we report results from regressing effort provision on returns to effort beliefs, which is instrumented by a full-interaction between dummies indicating quiz difficulty and payoff feedback treatments (Easy-Mixed 1, Difficult-Mixed 1, Easy-Mixed 2, Difficult-Mixed 2). The results reported in Columns 1 and 3 show that a 10 percentage points increase in returns to effort beliefs leads to 1 additionally solved decoding task ($p < 0.001$) and 14 additional seconds of work time ($p = 0.010$). Next, we examine how effort provision responds to misguidedness, using the same set of instruments. Recall that negative misguidedness is theorized to discourage effort provision, and positive misguidedness is theorized to encourage it. The results reported in Columns 2 and 4 show that a 10 percentage points increase in the degree of misguided learning leads to 1.1 additionally solved decoding task ($p = 0.001$) and 14 additional seconds of work time ($p = 0.017$). Overall, the results in Table 3 show that returns to effort beliefs and the degree of misguided learning have a positive and causal effect on effort provision, as predicted in hypothesis 2.

Next, we quantify the distortions to effort provision due to misguided learning that arise as a result of initial biases in priors. We calculate deviations from implied optimal effort provision levels by using the coefficient estimates for returns to effort beliefs in Columns 1 and

3 of Table 3 to predict the level of effort provision participants would be expected to provide if they held unbiased performance priors. We call the difference between a participant's effort provision and the predicted level of effort provision *misguided effort*. Figure 7 plots the means and 95% confidence intervals of misguided effort, separately for participants who formed positively ($\Delta\theta > 0$) and negatively ($\Delta\theta \leq 0$) misguided returns to effort beliefs.

Panel A in Figure 7 shows the results for the number of solved decoding tasks effort metric. We find that positively misguided participants solve on average 3 more decoding tasks than the predicted optimal level of solved decoding tasks (Wilcoxon signed-rank test, $p < 0.001$). The opposite pattern emerges for negatively misguided participants, who solve on average 2.5 less decoding tasks than the predicted optimal level of solved decoding tasks (Wilcoxon signed-rank test, $p < 0.001$). Panel B in Figure 7 shows the results for the time spent on the decoding tasks effort metric. We find that positively misguided participants work on average 41.1 seconds longer than the predicted optimal work time (Wilcoxon signed-rank test, $p < 0.001$). Negatively misguided participants, on the other hand, work on average 34.4 seconds shorter than the predicted optimal work time (Wilcoxon signed-rank test, $p < 0.001$).

Panel A: Misguided Solved Decoding Tasks



Panel B: Misguided Decoding Time

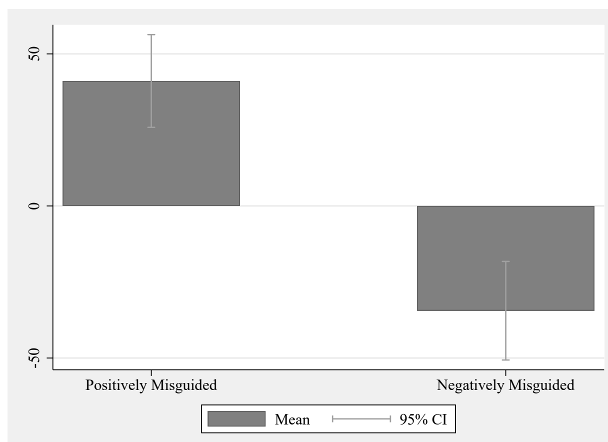


Figure 7: Misguided Effort, Split by the Direction of Misguided Learning

Across all participants, individuals deviate from the effort they would be expected to provide if they had accurate priors by 9.3 solved decoding tasks (in absolute terms) and by 153 seconds of work time. Recall that participants solved 12.8 tasks and worked for 216 seconds on average. In this context, the extent of misguided effort is economically significant.

4 Discussion

We examined the impact of initial beliefs about one's performance on inferences about external fundamentals that causally affect economic actions. Our results suggest that miscalibrated initial beliefs can lead to sub-optimal effort provision. Specifically, we show that overconfident individuals who experience bad initial labor market outcomes form more pessimistic beliefs about the returns to effort, and therefore put less effort into the subsequent real-effort task, compared to overconfident individuals who experience good initial outcomes. Consequently, overconfident individuals who experience bad initial labor market outcomes end up with substantially lower payoffs, compared to overconfident individuals who experience good initial outcomes. Underconfident individuals also learn misguidedly and adjust their efforts accordingly, but in the opposite direction.

The implications of this evidence are broad, as the inference about one's performance and the external factors that impact one's progress is complicated by many dimensions in the real-world. First, individuals rarely work alone. Therefore, how their input interacts with the input of other team members, and how their performance is evaluated in a team context are important external factors that influence individual outcomes. Second, although a meritocratic society has been idealized, the fruits of labor are rarely determined only by merit. Factors such as gender, race, economic background, social network, etc. may influence the returns to effort and education. Third, it may be unclear whether one's performance feedback is absolute or compared to peers, and what comparative performance level could be achieved with additional effort, as it requires knowledge about peers' abilities and effort decisions. In sum, the real-world presents many complex reasons why individuals cannot easily identify returns to effort from how much their efforts are rewarded at any given circumstance. We focus only on learning about whether the returns to effort are positive or zero in a stylized context. We hope our experimental paradigm and results spur more research into the nuances of the impact of miscalibrated initial beliefs on effort provision across these different domains.

Our focus has been to provide evidence for the causal impact of miscalibrated priors on economically relevant actions when the environment fosters misguided learning about external fundamentals and hinders learning about one's potential. Our design purposefully eliminates learning about one's ability to rule out a confounding impact on effort provision that might arise as a result. However, we remain interested in how beliefs would evolve in the long-run and in settings where feedback might also contain information about ability. It is conceivable that initial belief biases trigger a detrimental cycle of misguided learning, leading to downward or upward spirals in one's beliefs about external fundamentals that is triggered by initial belief biases and experiences in life. For instance, overconfident individuals who experience bad initial labor market outcomes may become more pessimistic about the economic environment and consequentially work less hard, which reinforces the propensity to experience bad labor market outcomes in the future. Such long-run effects are especially interesting to study in field settings. Having established that misguided learning matters for effort provision, we hope that future work examines the long-term consequences of initial belief biases in the field.

References

Arkin, R. M. and G. M. Maruyama (1979): "Attribution, affect, and college exam performance." *Journal of Educational Psychology*, 71, 85.

Barber, B. M. and T. Odean (2001): "Boys will be boys: Gender, overconfidence, and common stock investment," *The quarterly journal of economics*, 116, 261–292.

Barron, K. and C. Gravert (2022): "Confidence and career choices: An experiment," *The Scandinavian Journal of Economics*, 124, 35–68.

Bénabou, R. and J. Tirole (2005): "Self-confidence and personal motivation," *Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions*, 19–57.

Brunnermeier, M. K. and J. A. Parker (2005): "Optimal expectations," *American Economic Review*, 95, 1092–1118.

Camerer, C. and D. Lovallo (1999): "Overconfidence and excess entry: An experimental approach," *American economic review*, 89, 306–318.

Chen, S. and H. Schildberg-Hörisch (2019): "Looking at the bright side: The motivational value of confidence," *European Economic Review*, 120, 103302.

Civelli, A., C. Deck, et al. (2018): "A Flexible and Customizable Method for Assessing Cognitive Abilities," *Review of Behavioral Economics*, 5, 123–147.

Coutts, A., L. Gerhards, and Z. Murad (2022): "Who to blame? Self-serving attribution bias with multi-dimensional uncertainty," Tech. rep., working paper.

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

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2022): "Estimating social preferences and gift exchange at work," *American Economic Review*, 112, 1038–1074.

Erkal, N., L. Gangadharan, and B. H. Koh (2018): "Monetary and non-monetary incentives in real-effort tournaments," *European Economic Review*, 101, 528–545.

Goerg, S. J., S. Kube, and J. Radbruch (2019): "The effectiveness of incentive schemes in the presence of implicit effort costs," *Management Science*, 65, 4063–4078.

Goette, L. and M. Kozakiewicz (2022): "Experimental evidence on misguided learning," Tech. rep.

Heidhues, P., B. Kőszegi, and P. Strack (2018): "Unrealistic expectations and misguided learning," *Econometrica*, 86, 1159–1214.

Hestermann, N. and Y. Le Yaouanq (2021): "Experimentation with self-serving attribution biases," *American Economic Journal: Microeconomics*, 13, 198–237.

Lau, R. R. and D. Russell (1980): "Attributions in the sports pages." *Journal of personality and social psychology*, 39, 29.

Lobeck, M. (2023): "Motivating beliefs in a just world," *Available at SSRN 4369242*.

Malmendier, U. and G. Tate (2005): "CEO overconfidence and corporate investment," *The journal of finance*, 60, 2661–2700.

Malmendier, U. and T. Taylor (2015): "On the verges of overconfidence," *Journal of Economic Perspectives*, 29, 3–8.

Marray, K., N. Krishna, and J. Tang (2020): "How do expectations affect learning about fundamentals? some experimental evidence," *Working paper*.

Mezulis, A. H., L. Y. Abramson, J. S. Hyde, and B. L. Hankin (2004): "Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias." *Psychological bulletin*, 130, 711.

Moore, D. A. and P. J. Healy (2008): "The trouble with overconfidence." *Psychological review*, 115, 502.

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

Ross, M. and F. Sicoly (1979): "Egocentric biases in availability and attribution." *Journal of personality and social psychology*, 37, 322.

Schwardmann, P., E. Tripodi, and J. J. Van der Weele (2022): "Self-persuasion: Evidence from field experiments at international debating competitions," *American Economic Review*, 112, 1118–1146.

Schwardmann, P. and J. Van der Weele (2019): "Deception and self-deception," *Nature human behaviour*, 3, 1055–1061.

Solda, A., C. Ke, L. Page, and W. Von Hippel (2020): "Strategically delusional," *Experimental Economics*, 23, 604–631.

Van den Steen, E. (2004): "Rational overoptimism (and other biases)," *American Economic Review*, 94, 1141–1151.

Appendices

A Balance Across Treatments

Columns 1 and 2 of Table A1 show the means of observable characteristics across quiz difficulty and payoff feedback treatments. Columns 3 and 4 of Table A1 show the magnitudes of the differences and the corresponding p-values. All reported p-values are at least greater than 0.34 and confirm balance in observable characteristics across quiz difficulty and payoff feedback treatments.

Table A1: Confirming Balance

	(1) Easy (N=505)	(2) Difficult (N=499)	(3) Difference	(4) p-value
Female	0.51	0.48	0.03	0.35
Common Questions Score	4.86	4.94	-0.07	0.34
	Mixed 1 (N=502)	Mixed 2 (N=502)	Difference	p-value
Female	0.49	0.49	0	1
Logic Score	7.99	7.90	0.09	0.85
Common Questions Score	4.95	4.85	-0.09	0.48
Performance Priors	59.37	58.80	0.56	0.70
Prior Bias	8.56	8.59	-0.03	0.89

Notes:
For the comparisons of female, the p-values are based on Fischer's exact tests. For all other comparisons, the p-values are based on Wilcoxon rank-sum tests.

B Comparison to Bayesian Benchmark

In the following analysis, we investigate how participants' updating behavior of performance priors into returns to effort beliefs follows the Bayesian prediction. Overall, participants' returns to effort beliefs are systematically correlated with Bayesian returns to effort beliefs ($\rho = 0.489$). To quantify the divergence from Bayesian predictions, we compute the individual *updating bias* for each participant. Updating bias is defined as the difference between a participant's returns to effort belief and the Bayesian returns to effort belief, given the participant's performance prior. The distribution of updating bias is illustrated in Figure A1. It shows that about 1/3 of the participants deviate by less than 5 percentage points and 2/3 of the participants deviate by less than 20 percentage points from the Bayesian prediction. Participants who demonstrate a substantial updating bias, deviating 50 percentage points from the Bayesian prediction, comprise 6.4% of the sample.

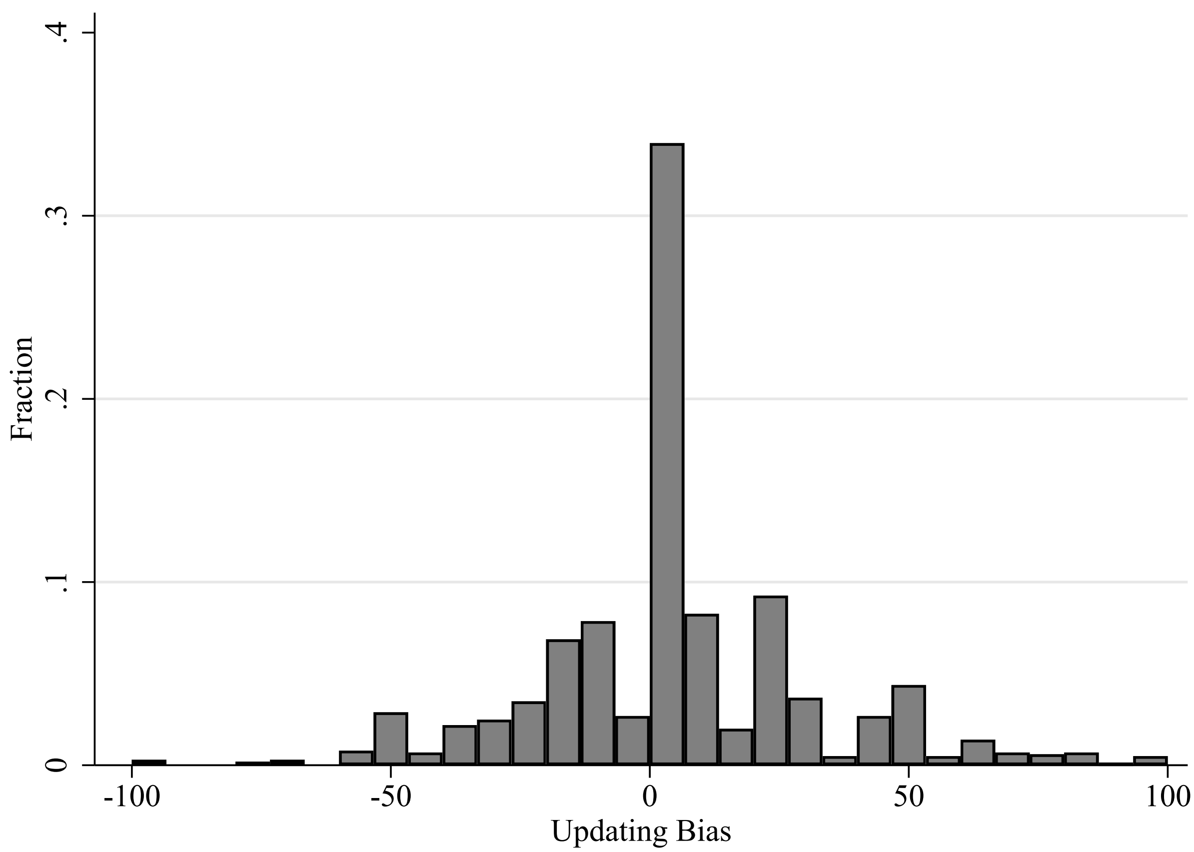


Figure A1: Distribution of Participants' Updating Bias

On average, the updating bias is 3.7 and therefore significantly higher than 0 (Wilcoxon signed-rank test, $p < 0.001$). This result implies that participants form systematically higher beliefs about the returns to effort than the Bayesian prediction. Based on the literature on motivated beliefs, we propose two potential explanations for this result. First, participants may form higher beliefs about the returns to effort in order to motivate themselves to work harder in period 2 as suggested by Bénabou and Tirole (2005) and Lobeck (2023). Second, participants may form higher beliefs about the returns to effort to derive anticipatory utility from apparently higher expected income streams in period 2 as proposed by Brunnermeier and Parker (2005). We remain agnostic about the exact behavioral foundation, but the observed over-optimistic returns to effort beliefs suggest that the joint inference about one's own ability and an external fundamental provides a scope for motivated belief distortions as suggested by Coutts et al. (2022).

C Experimental Instructions

Welcome to our study! As a reminder, you must complete this survey on a desktop computer. No mobile devices are allowed. Please click next to continue.

Consent Form

Principal Investigator: Yesim Orhun (University of Michigan)
Co-investigator: Christoph Drobner (Technical University Munich)

You are invited to participate in a research study about decision-making. If you agree to be part of the research study, you will be asked to make decisions on your computer. Data collected during the experiment will be linked only to your Prolific ID, and in no way will it be linked to your name. Therefore, your behavior during this study is completely anonymous.

Benefits of the research:
You may not receive any personal benefits from being in this study. However, others may benefit from the knowledge gained from this study.

Risks and discomforts:
This study involves behavioral tasks and does not pose more than minimal risks to you physically, psychologically, and legally.

Compensation:
You will receive a completion fee of $2. In addition, you may receive additional payments based on the decisions made by you and the other participants during the study. Further details will be given in the instructions when the study begins.

Participating in this study is completely voluntary. Even if you decide to participate now, you may change your mind and stop at any time. If you have questions about this research study, please contact Christoph Drobner (christoph.drobner@tum.de).

As part of their review, the University of Michigan Institutional Review Board Health Sciences and Behavioral Sciences has determined that this study is no more than minimal risk and exempt from on-going IRB oversight.

Please click the arrow button if you attest to being at least 18 years old and agree to take part in the study. Please close the browser window if you are not willing to take part in the study.

Welcome to this study!

For participating in this study, you may earn payments depending on your decisions. Hence, please read the following instructions carefully.

There are two parts to today's study: Part A and Part B. You will receive detailed instructions for each part before you participate in them.

You will be paid for one of the two parts, chosen at random by the computer. This means that you should consider the decisions you make in each part carefully.

All the decisions you make in this study will be anonymous.

Finally, please note that this is a no-deception study. All the instructions and information you will receive in this study are true and accurate.

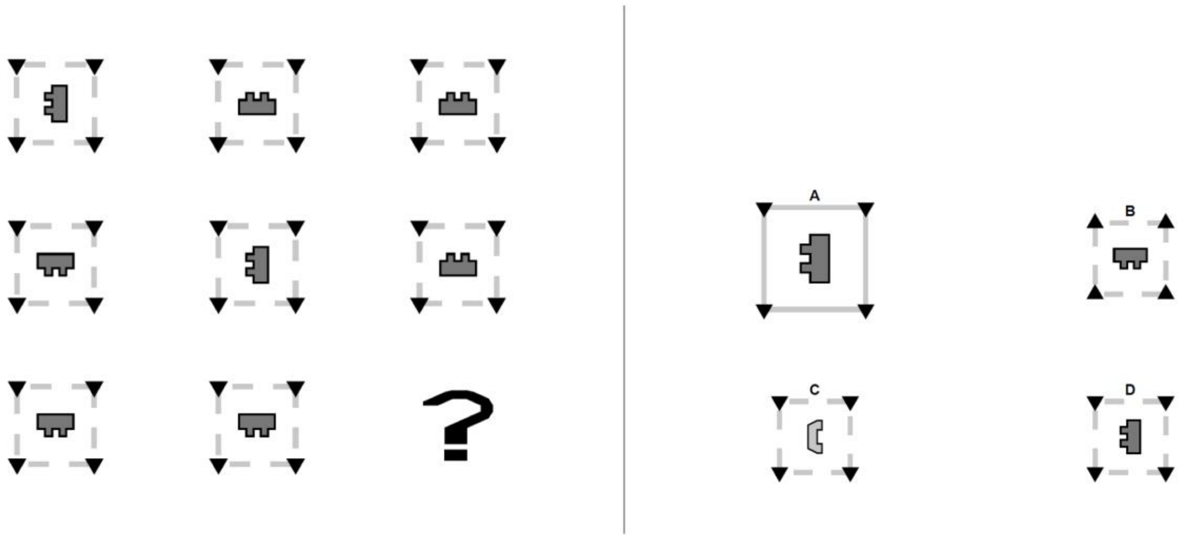
Part A

In Part A, you will solve a logic quiz with 12 questions. You will have 20 seconds to solve each question.

Your expected earnings will increase in the number of correctly solved questions in the logic quiz. This means, the higher your number of correctly solved questions, the more likely you will earn a higher payoff from the logic quiz (the range is $0-$4). So please do your best.

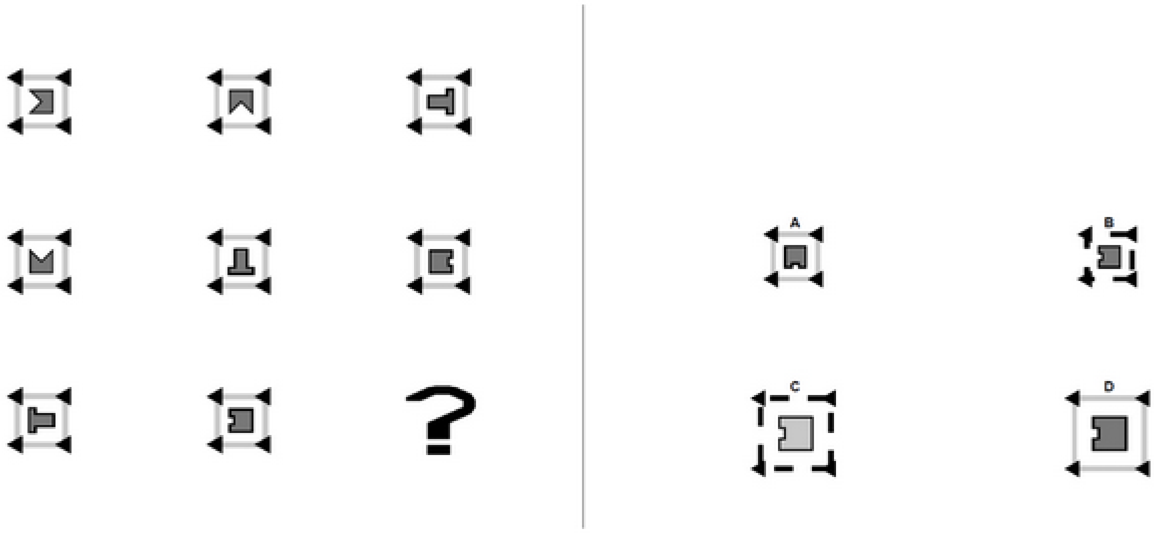
On the next page, we present a sample question of the logic quiz. The questions ask you to select the most logical option (A, B, C or D) that belongs to the ? in the pattern you see.

Sample Question



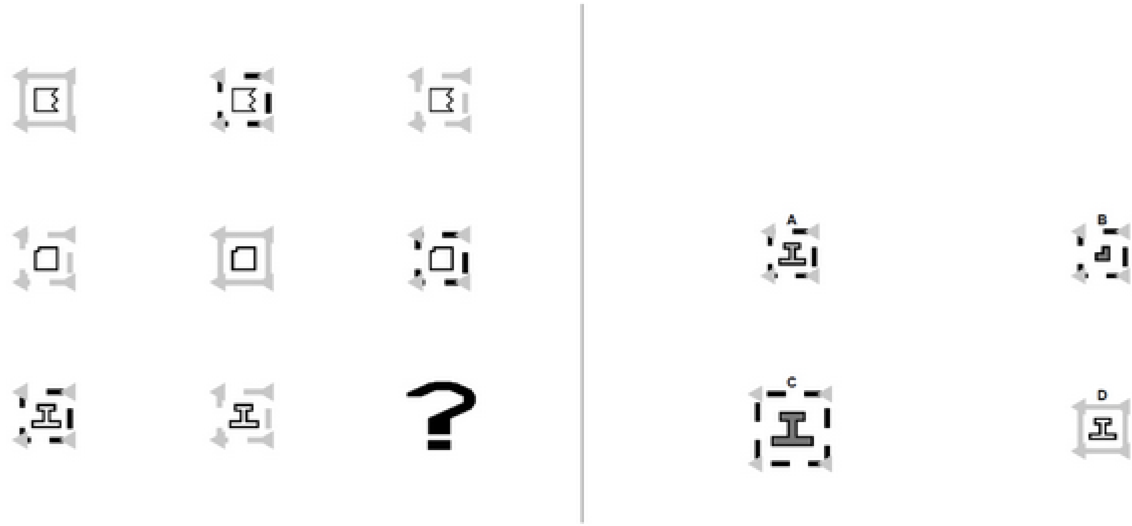
The correct answer to the sample question is Option D. You will see questions of varying difficulty. If a question feels hard to solve, it is probably a difficult question for everyone. Do your best at all times! The better you perform, the higher your payoff is expected to be. Please click the arrow button if you are ready to start the quiz.

Question (1/12)



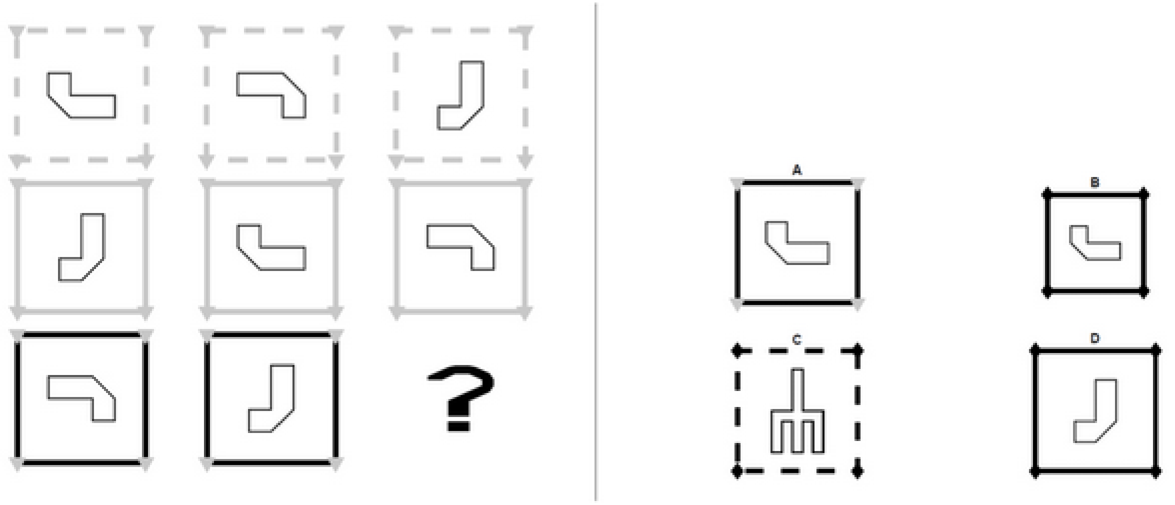
A	B	C	D
○	○	○	○

Question (2/12)



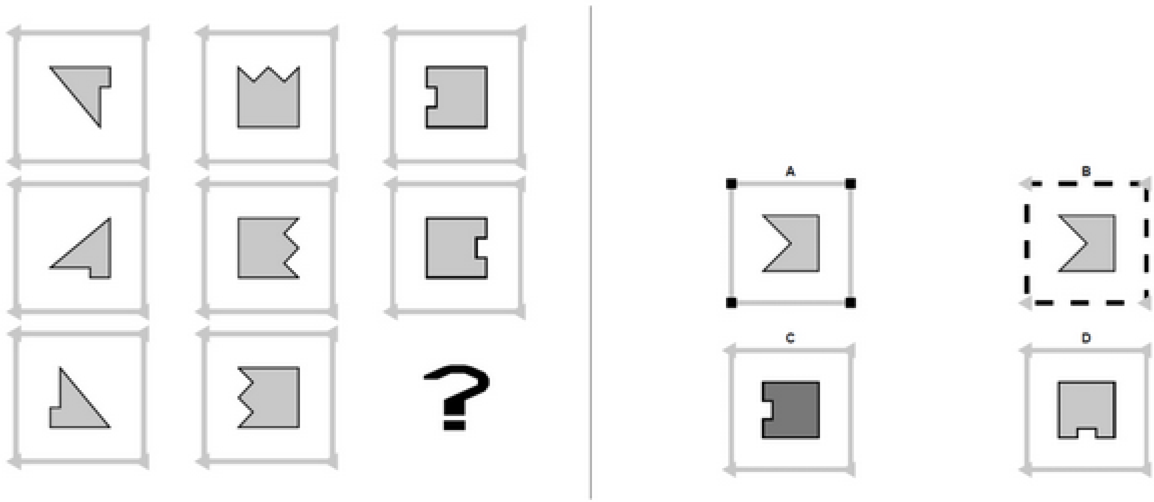
A	B	C	D
◯	◯	◯	◯

Question (3/12)



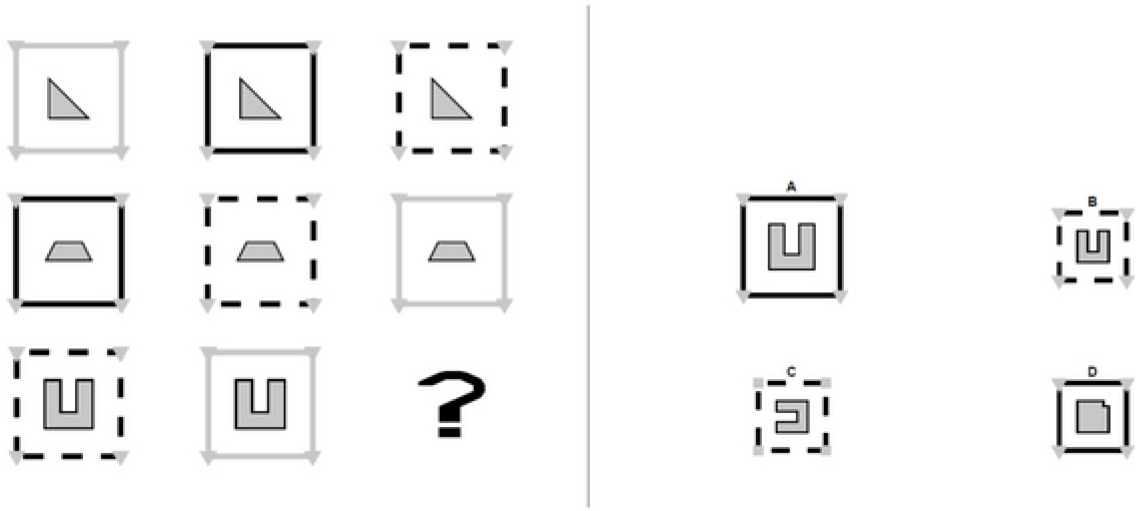
A	B	C	D
○	○	○	○

Question (4/12)



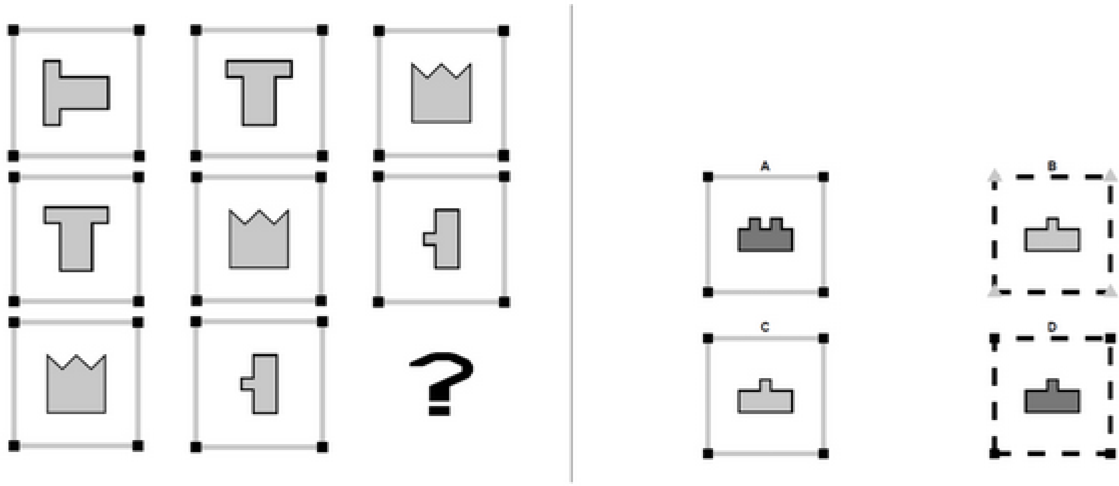
A	B	C	D
◯	◯	◯	◯

Question (5/12)



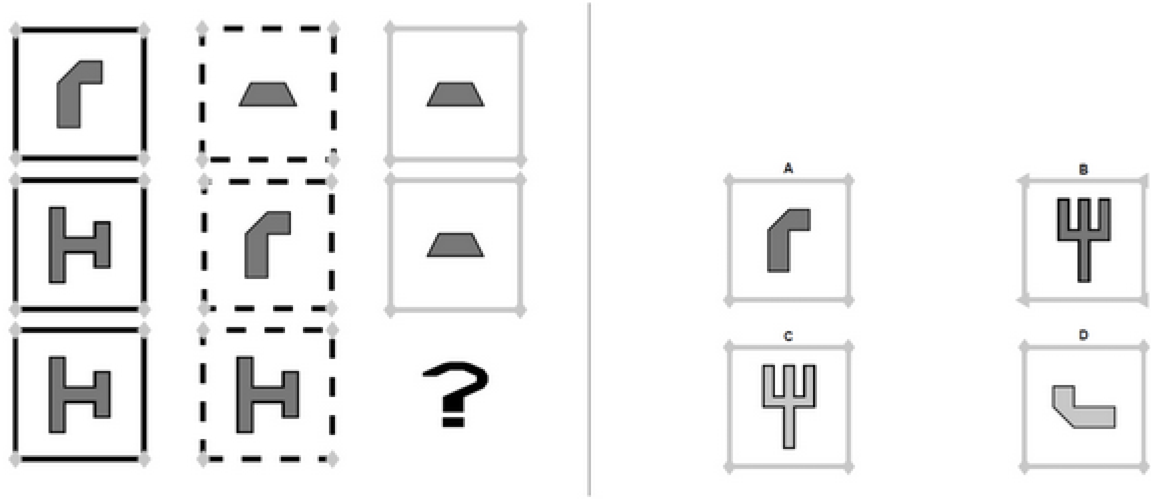
A ◯ | B ◯ | C ◯ | D ◯

Question (6/12)



A	B	C	D
○	○	○	○

Question (7/12)

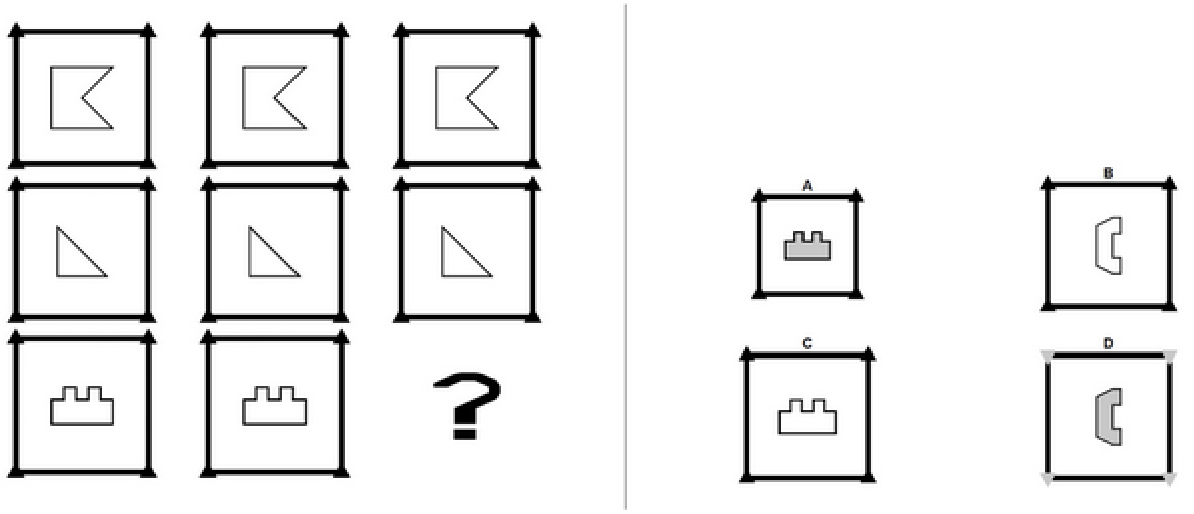


A	B	C	D
◯	◯	◯	◯

QUESTIONS 8-12 VARY BETWEEN QUIZ DIFFICULTY TREATMENTS

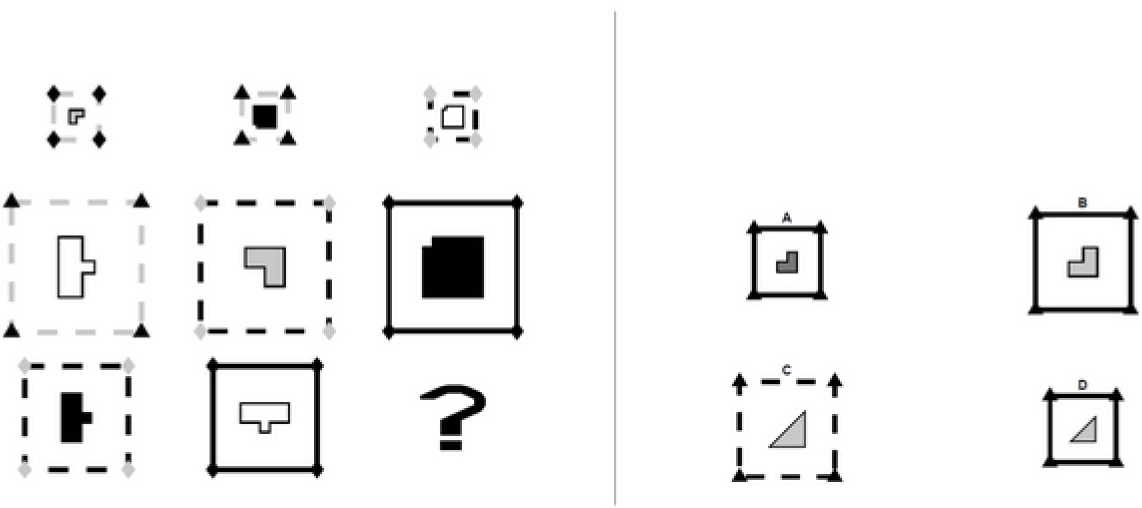
Question (8/12)

EASY



A	B	C	D

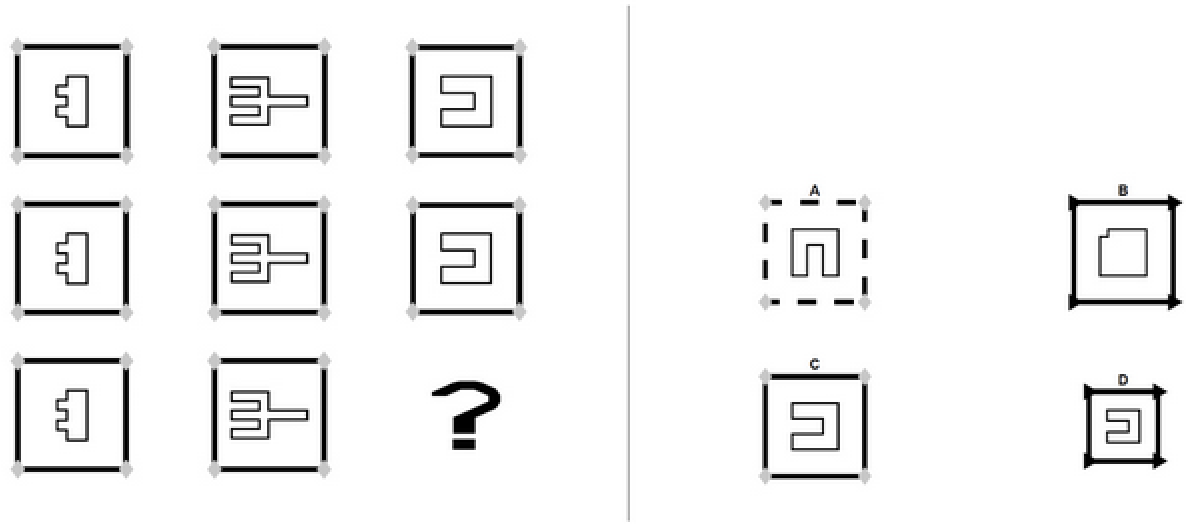
DIFFICULT



A	B	C	D

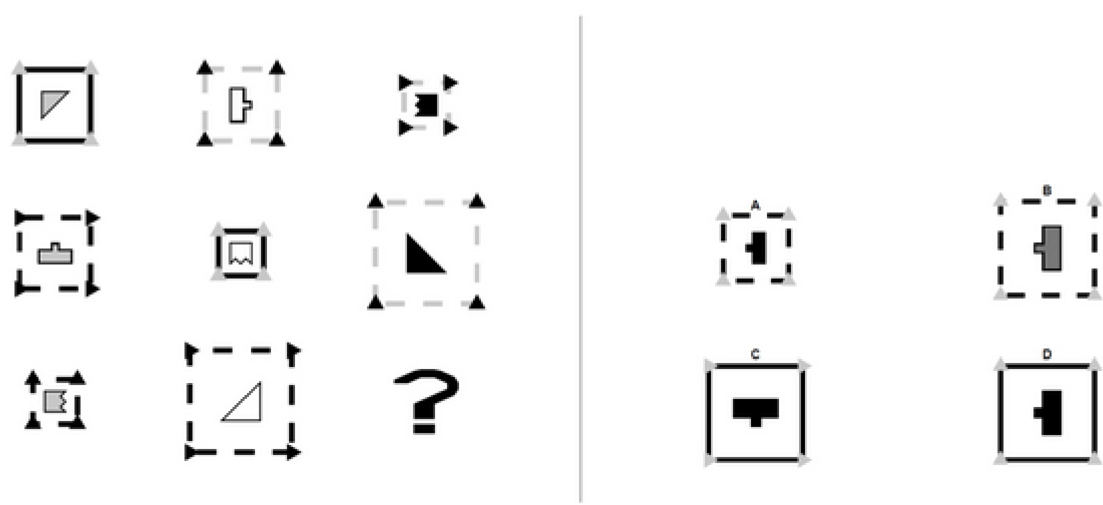
Question (9/12)

EASY



A	B	C	D
○	○	○	○

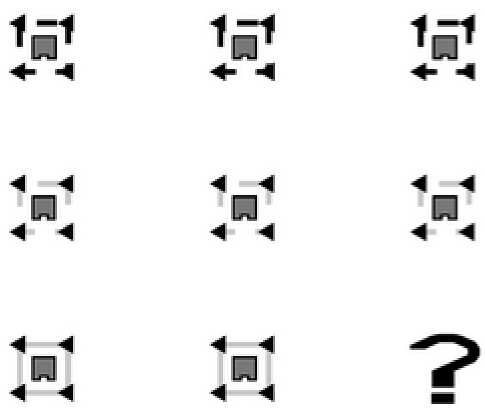
DIFFICULT

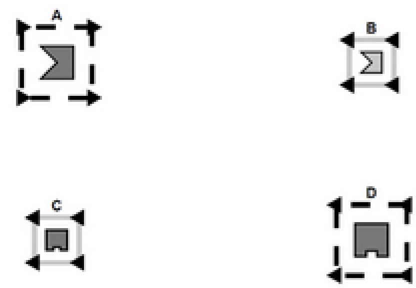


A	B	C	D
○	○	○	○

Question (10/12)

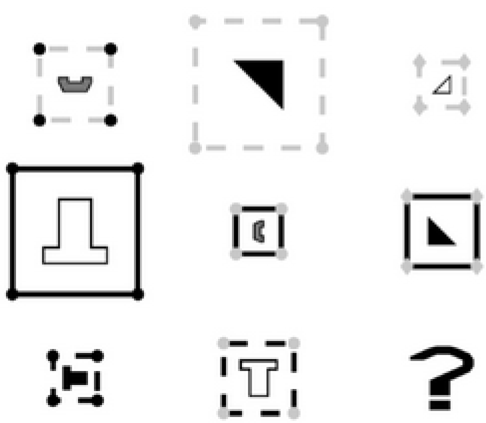
EASY

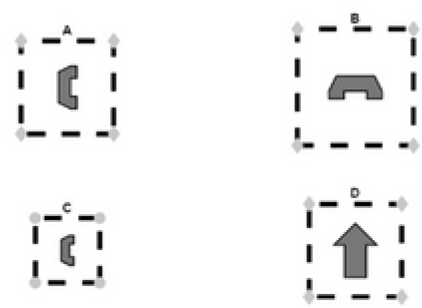




A	B	C	D
◯	◯	◯	◯

DIFFICULT

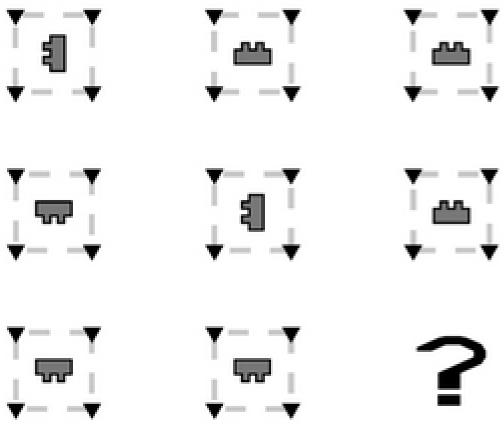


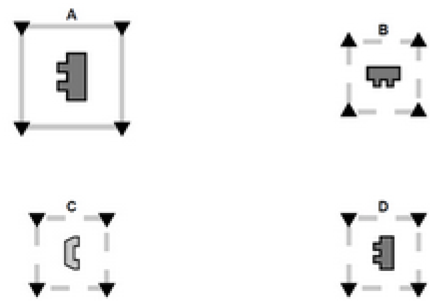


A	B	C	D
◯	◯	◯	◯

Question (11/12)

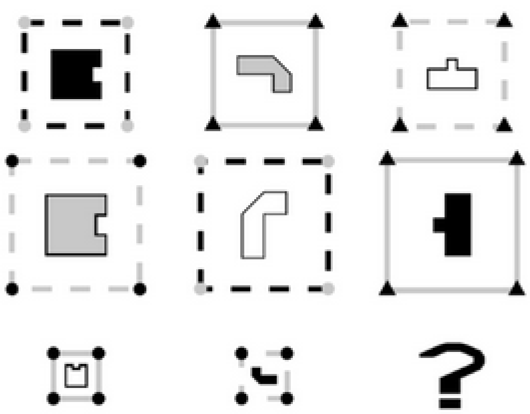
EASY

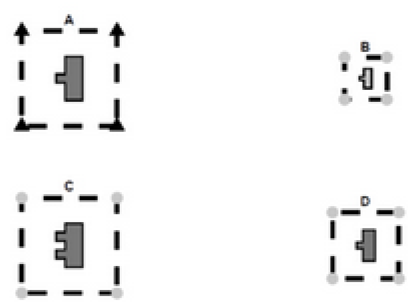




A	B	C	D
◯	◯	◯	◯

DIFFICULT

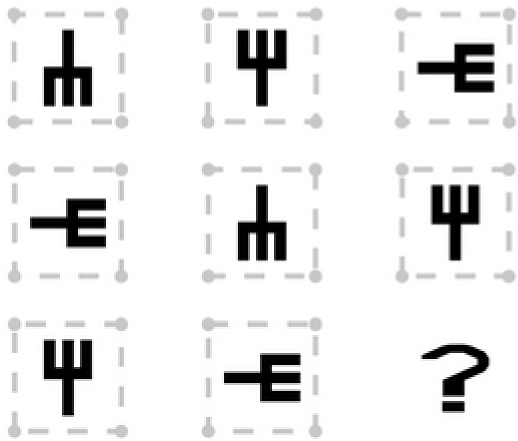


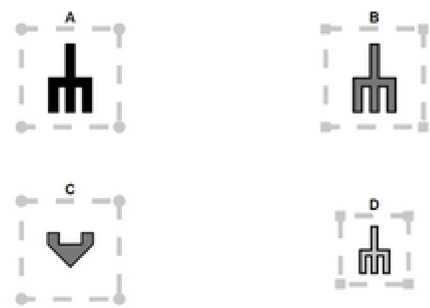


A	B	C	D
◯	◯	◯	◯

Question (12/12)

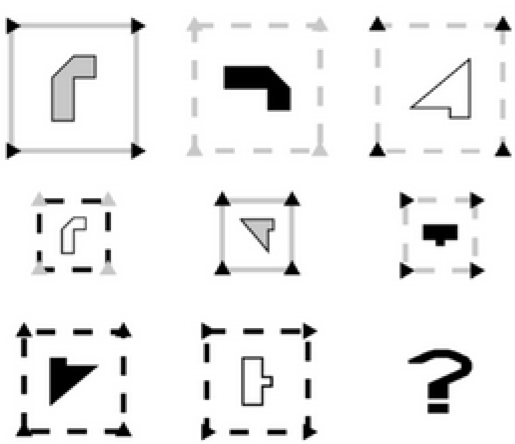
EASY

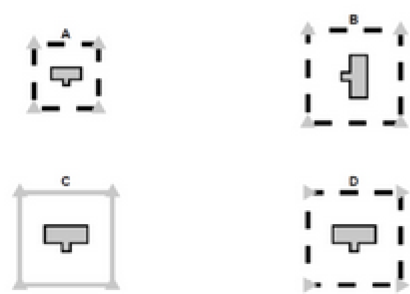




A	B	C	D
◯	◯	◯	◯

DIFFICULT





A	B	C	D
◯	◯	◯	◯

You have solved **(8/12)** questions of the logic quiz correctly.

The program compared your score in the logic quiz with other participants in this study. Your comparison group consists of a group of 4 participants (including you). We want you to estimate your relative performance in the logic quiz among your comparison group. We ask you to guess how likely it is that you were ranked in the top half or the bottom half among your comparison group.

We will reward the accuracy of your guess by using a payment rule that secures the highest chance of winning **$1** when you provide your most-accurate guess.

Please provide your best guess: What is the percent chance that you ranked in the top half or the bottom half among your comparison group? (must add up to 100)
Percent chance that I ranked in the **top half** : _______
Percent chance that I ranked in the **bottom half** : _______
Total : ________

Please read the instructions that follow very carefully. You will be asked to answer 2 comprehension questions.

Please proceed.

Payoffs for Logic Quiz

You will receive two payoffs for participating in the logic quiz. One payoff is provided by Evaluator 1, and the other by Evaluator 2. Exactly one of the two evaluators is a Performance Evaluator and exactly one of the two evaluators is a Random Evaluator.

- **Performance Evaluator**: The performance evaluator's payoff is determined based on your performance in the logic quiz compared to a group of 4 participants, including yourself. If your performance ranked in the top half of the group, the performance evaluator pays you $2. On the other hand, if your performance ranked in the bottom half, the performance evaluator does not provide any payment, resulting in a payoff of $0.
- **Random Evaluator**: The random evaluator determines your payoff by tossing a coin. If the coin toss results in heads, the random evaluator pays you $2. If the coin toss results in tails, the random evaluator does not provide any payment, resulting in a payoff of $0.

You will see the payoffs of Evaluator 1 and Evaluator 2. However, you will not know which of the two evaluators is the Performance or Random Evaluator.

(When ready, click next to see comprehension questions about these instructions)

Which of the following statements about the payoffs for the logic quiz are true? (please select all true statements)

- Exactly one of the two evaluators will be the Performance Evaluator and exactly one of the two evaluators will be the Random Evaluator.
- It is possible that both Evaluator 1 and Evaluator 2 are the Performance Evaluator.
- It is possible that both Evaluator 1 and Evaluator 2 are the Random Evaluator.
- One evaluator determines my payoff by a coin toss and one evaluator determines my payoff by my relative performance in the logic quiz.

Which of the following statements about the payoff from the Performance Evaluator are true? (please select all true statements)

- The Performance Evaluator pays me $2 if my performance in the logic quiz ranks in the top half of my comparison group with 4 participants.
- The Performance Evaluator pays me $2 if my performance in the logic quiz ranks in the bottom half of my comparison group with 4 participants.
- The Performance Evaluator pays me $0 if my performance in the logic quiz ranks in the top half of my comparison group with 4 participants.
- The Performance Evaluator pays me $0 if my performance in the logic quiz ranks in the
- bottom half of my comparison group with 4 participants.

Payoffs for Logic Quiz

You have received the following payoffs from the evaluators:

THE FOLLOWING INSTRUCTIONS VARY BETWEEN PAYOFF FEEDBACK TREATMENTS

BOTH HIGH
Evaluator 1: **$2**
Evaluator 2: **$2**

BOTH LOW
Evaluator 1: **$0**
Evaluator 2: **$0**

MIXED 1
Evaluator 1: **$2**
Evaluator 2: **$0**

MIXED 2
Evaluator 1: **$0**
Evaluator 2: **$2**

The sum determines your payment for participating in the logic quiz.

Thank you. You are done with Part A. You are now moving onto Part B.

Part B

In Part B, you will be given a series of decoding tasks.

One such decoding task is illustrated below. Your task is to decode text from a number and enter the answer into an input field. In this example, the correct answer is WKBLE. This solution is achieved by looking up the corresponding letter for each number in the panel. Solving decoding tasks correctly requires attention, patience and effort.

W	U	H	J	P	L	E	B	Z	K
9	5	2	4	6	3	7	8	0	1

91837

Please make sure you understand how to decode the number before you proceed. You will not see these instructions again.
The compensation details follow.

Payoffs for Part B (READ CAREFULLY)

In **Part B**, you will only be paid by the same **Evaluator 1** from Part A.

If Evaluator 1 is the Performance Evaluator, your earnings from Part B will again depend on your performance. This time, the performance evaluator pays you **10 c**ents for each decoding task you solve correctly independent of the performance of other participants. The more decoding tasks you answer correctly, the more money you can make.

If Evaluator 1 is the Random Evaluator, regardless of how much you work, you will receive **no payoff**.

Recall that in Part A, Performance Evaluator paid $2 if your performance in the logic quiz was in the top half, and $0 if not. Random Evaluator paid $2 or $0 randomly.

THE FOLLOWING INSTRUCTIONS VARY BETWEEN PAYOFF FEEDBACK TREATMENTS

MIXED 1
You received one payment from each of them: **Evaluator 1 paid you $2** and **Evaluator 2 paid you $0.**

MIXED 2
You received one payment from each of them: **Evaluator 1 paid you $0** and **Evaluator 2 paid you $2.**

Given that you expected to **rank in the top half with 60% chance**, you may have some idea about which type of evaluator Evaluator 1 is.

Please provide your best guess: What is the percent chance that Evaluator 1 is the Performance or the Random Evaluator? (must add up to 100)
Percent chance that Evaluator 1 is the **Performance Evaluator** : _______
Percent chance that Evaluator 1 is the **Random Evaluator** : _______
Total : ________

Recall that the Performance Evaluator pays you 10 cents per correctly solved task, and the Random Evaluator pays you 0 cents regardless of how many tasks you solve. The maximum number of decoding tasks you can solve is 25.

According to your answer on the previous page, you think being evaluated by the Performance Evaluator is more likely.

You can **decide how much you want to work** after completing the first task. You can solve as few as 1 or as many as 25 decoding tasks.
After each task, you will be given the chance to decide whether you want to continue or stop working. If you stop, you will be forwarded to the end of the survey. Please click the arrow button when you're ready to begin working.

Please enter text decoded from the number. This is achieved by looking up the corresponding letter for each number.

S	X	A	J	Z	L	N	G	M	F
8	3	5	2	9	0	4	6	7	1
38921									

Thank you. Recall that you think being evaluated by the Performance Evaluator is more likely. The Performance Evaluator pays you 10 cents, and the Random Evaluator pays you 0 cents for each correctly solved decoding task.

Do you want to continue or stop working and get to the end of the survey?

- Continue working
- Stop working and get to the end of the survey

Thank you for your time spent taking this survey. The program will calculate your bonus payment and credit it to your Prolific account within five business days.
Please click the button below to be redirected back to Prolific and register your submission.