Misguided Effort*

Christoph Drobner†
A. Yeşim Orhun‡

First Draft: January 23, 2024

Abstract

We experimentally study how miscalibrated prior beliefs about one's own ability affect effort provision through misguided inferences about returns to effort. We demonstrate 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 their own ability and external fundamentals. Crucially, we establish that misguided inferences lead to suboptimal future effort provision. These findings provide the first causal empirical support for a theorized effect of miscalibrated prior beliefs on economic actions that operates through misguided inferences about the economic environment.

Keywords: Beliefs, Overconfidence, Underconfidence, Misguided Inference, 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

Our learning environments are complex. We often need to learn about ourselves and our economic environments from outcomes that are determined by both. For example, consider a student who gets a low grade on his math exam. Was it because he did not study much, because he is not good at math, or because the teacher does not like him? It is hard for the student to judge the contribution of each of these potential factors. His priors about his math ability, the returns to studying, and whether his teacher is biased are likely to shape his inferences. These inferences, in turn, may shape how much effort he puts into math going forward. Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) model this inference problem and show that miscalibrated priors about one determinant (e.g., about one's own ability) can lead to misguided inferences about other determinants (e.g. about an external fundamental) in learning environments where outcomes are multiply-determined.[1] Importantly, their work suggests that misguided inference may lead to suboptimal economic choices.

In this paper, we experimentally study how miscalibrated prior beliefs about one's ability causally affects effort provision through misguided inferences about the returns to effort. In the context of a labor market, participants work in an environment that either pays them based on chance or based on their relative performance. Workers observe their earnings, form beliefs about the returns to effort and decide on their effort level in a subsequent task. Our experimental design introduces a novel feedback treatment that fosters misguided inferences about the returns to effort, while preventing any learning regarding personal ability. This design element aligns our experiment closely with the theoretical framework presented in Heidhues et al. (2018) and is key for the identification of our findings. The results demonstrate that miscalibration about one's own ability causally affects effort provision through misguided inferences about returns to effort, corroborating the theory of Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021).

To demonstrate the causal link between initial belief biases and their impact on effort provision through misguided inference, our experiment utilizes orthogonal interventions to manipulate both the extent of bias in prior beliefs and the direction of the resulting misguided inference, being careful not to introduce a confounding path between the treatment that manipulates misguided inference and the effort provision. The experiment, illustrated by Figure 1 in Section 2.1, consists of 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

[1] Misspecified learning, more broadly, can arise due to incorrect priors (e.g. Ba and Gindin, 2023; Fudenberg et al., 2017, 2021; Nyarko, 1991) or incorrect mental models of the data generating process (e.g. Bohren and Hauser, 2021; Esponda and Pouzo, 2016; Frick et al., 2020, 2022)

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. Consequently, participants receive one of four types of payoff feedback based on their logic quiz performance, the outcome of the coin toss 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 receive no information about their ability and cannot separately identify the contribution of their logic quiz performance and external luck, paralleling situations that Heidhues et al. (2018) study.

In the second period, participants who received mixed payoff feedback work on real effort tasks. Crucially, we inform participants that their earnings in the second period depend on 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 *misguided inference* as the difference between these subjective returns to effort beliefs and objective inferences Bayesian participants would have constructed if they held accurate performance priors. Then, participants solve up to 25 decoding tasks. Their effort provision is a choice: they can choose to stop working at any time. Section 2.1 provides further details about the experimental design and protocol.

We predict that individuals with overconfidence attribute a high payoff from evaluator 1 to the performance evaluator, while attributing the low payoff from evaluator 1 to the random evaluator. Conversely, we expect underconfident individuals to attribute a low payoff from evaluator 1 to the performance evaluator and the high payoff to the random evaluator. Thus, we anticipate that overconfident participants' inferences will be positively misguided if they are in the MIXED 1 group and negatively misguided if they are in the MIXED 2 group. Conversely, we expect underconfident participants to exhibit the opposite pattern of inference. Furthermore, we predict that workers whose inferences about returns to effort are positively misguided will tend to exert excessive effort in solving decoding tasks, whereas those whose inferences are negatively misguided will tend to exert insufficient effort. Section 2.3 provides a detailed account of these predictions.

In testing for the causal impact of misguided inference on effort provision, we rely on the payoff feedback treatment variation (Mixed 1 / Mixed 2) as an instrument for the direction of misguided inference. We carefully designed the payoff feedback treatment to introduce variation in the direction of misguided inference but not to introduce any other differences between these two groups that may impact effort provision (see Section 2.2 for details). 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 (e.g. Breza et al., 2018). Therefore, our payoff feedback treatment reveals no information about participants' relative rank in the logic quiz 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 misguided inference and effort provision across Mixed 1 and Mixed 2 groups. In addition to examining differences in the degree of misguided inference induced by the 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 inference about the returns to effort as a function of participants' prior bias and the experimental variation in payoff feedback, (2) the causal impact of this misguided inference on effort provision, and (3) the degree of misguided effort as a function of participants' prior bias. The results confirm all of our theoretical predictions. First, overconfident individuals' returns to effort inferences 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' inferences are negatively misguided by 14.5 percentage points in Mixed 1 and positively misguided by 19.1 percentage points in Mixed 2. Second, we show that misguided inference has a substantial impact on effort provision. Results suggest that a 10 percentage point increase in the degree of misguided inference leads to a causal increase in the second period effort equivalent to 11% of the median number of attempted decoding tasks and 8% of the median work time. Third, we confirm that initial prior bias dictates the degree of misguided effort. Overconfident individuals exert too much effort in period 2 when evaluator 1 provided the high payoff in period 1 while they exert too little effort when evaluator 1 provided the low payoff in period 1. The theorized opposite pattern emerges for underconfident individuals.

Overall, the amount of misguided effort is substantial. In absolute terms, participants deviate on average by 3.1 decoding tasks from the optimal number of decoding tasks they would have worked on if their inferences were not misguided, which corresponds to almost 1/3 of the median effort level.

Our findings offer the first empirical evidence of a causal relationship between prior biases and economic actions, mediated by misguided inference about the economic environment, 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 complements the burgeoning literature documenting misguided learning with optimistic beliefs about one's self (Galindo, 2023; Goette and Kozakiewicz, 2022; Marray et al., 2020; Özyılmaz, 2023) by examining the causal impact of misguided inference about the economic environment on economic actions. This literature examines the dynamic learning process that arises from initial overconfidence, illustrating how individuals consistently misinterpret outcomes and develop increasingly erroneous beliefs about the economic environment.[2] We abstract from this dynamic learning process to exclude the possibility that subjects' learning about their abilities confounds the causal impact of misguided inference on economic actions. Second, it extends 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 miscalibrated priors on economic behavior.[3] We spotlight the subtler, yet profound, effects that both overconfidence and underconfidence have on economic actions through misguided inferences about the economic environment. 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 as expected. We document a causal impact of beliefs on actions that is unconfounded by experimental design features and aligned with predictions of the standard

[2]A related literature in psychology has shown that individuals tend to attribute their achievements to personal merits and their failures to external factors. (see Mezulis et al., 2004, for a review). 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). In economics, evidence for other types of misattribution include individuals attributing their past experiences to inherent value rather than correctly recognizing that circumstances (Haggag et al., 2019, 2021) or initial expectations (Bushong and Gagnon-Bartsch, 2023) shaped their experience; individuals failing to neglect nuisance when learning from signals that contain diagnostic and nuisance components (Brownback and Kuhn, 2019; Erkal et al., 2022; Graeber, 2023).

[3]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)

model – higher returns to effort inferences cause individuals to provide more effort.

The practical implications of the presented evidence are broad, because in the real world, learning about one's ability and the external factors that impact one's progress is indeed complicated by the inference challenge that is at the center of our paper. 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 performance or effort. Factors such as gender, race, economic background, social network, etc. may influence the returns to effort and education. Third, when performance is evaluated comparatively to peers, it requires knowledge about peers' abilities and effort decisions. In sum, the real world presents many situations in which individuals cannot easily identify returns to effort from how much their efforts are rewarded at any given circumstance. We experimentally demonstrate how miscalibrated prior beliefs about one's own ability can lead to misguided inference about the returns to effort, and consequently suboptimal effort provision, when individuals cannot separately identify the effect of one's own ability and an external fundamental from observing labor market outcomes.

Methodologically, our experimental approach introduces a novel paradigm that shifts beliefs in one dimension while eliminating inferences in another. In our context, we exogenously move the direction of misguided inference about the returns to effort but we shut down the possibility that participants learn about their ability. This methodological innovation offers a versatile tool for future research exploring the dynamics of inference and choice in complex environments with multidimensional uncertainty. Moreover, our experimental design enables us to examine both the formation of biased beliefs and their influence on economic behavior in a cohesive manner. Previous studies examining biases in beliefs often implicitly assume that individuals rely on their biased beliefs when they make economically relevant decisions. However, it is plausible that people entertain one set of beliefs for the subjective utility derived from those beliefs, yet adopt another set of unbiased beliefs to inform their choices. Our experimental paradigm allows us to (1) address this potential criticism by showing that individuals rely on their biased beliefs when they make economically relevant decisions, and (2) provides a measure of the extent of harm miscalibrated priors can have through misguided inference and resulting actions.

2 Experimental Design and Theoretical Predictions

Theoretically, Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) show that when individuals cannot separately identify how much the quality of their input versus external fundamentals shape the feedback they receive, they will learn misguidedly about the external fundamental and provide sub-optimal levels of effort.[4] Our experiment tests this theorized causal chain of miscalibrated priors on effort provision in an environment where individuals form misguided beliefs about the returns to effort and do not learn about their abilities.

2.1 Overview

Figure 1 illustrates the experimental design and Appendix E provides all experimental instructions. The experiment consists of two periods. In both periods, participants work to solve tasks. At the end of the experiment, one period is randomly 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.[5] 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 in the form of two payoffs (s_1, s_2). One payoff is provided by an evaluator who pays participants based on their worker type: if the participant's performance ranks in the top half of the 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 evaluator types.

[4] 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.

[5] 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.

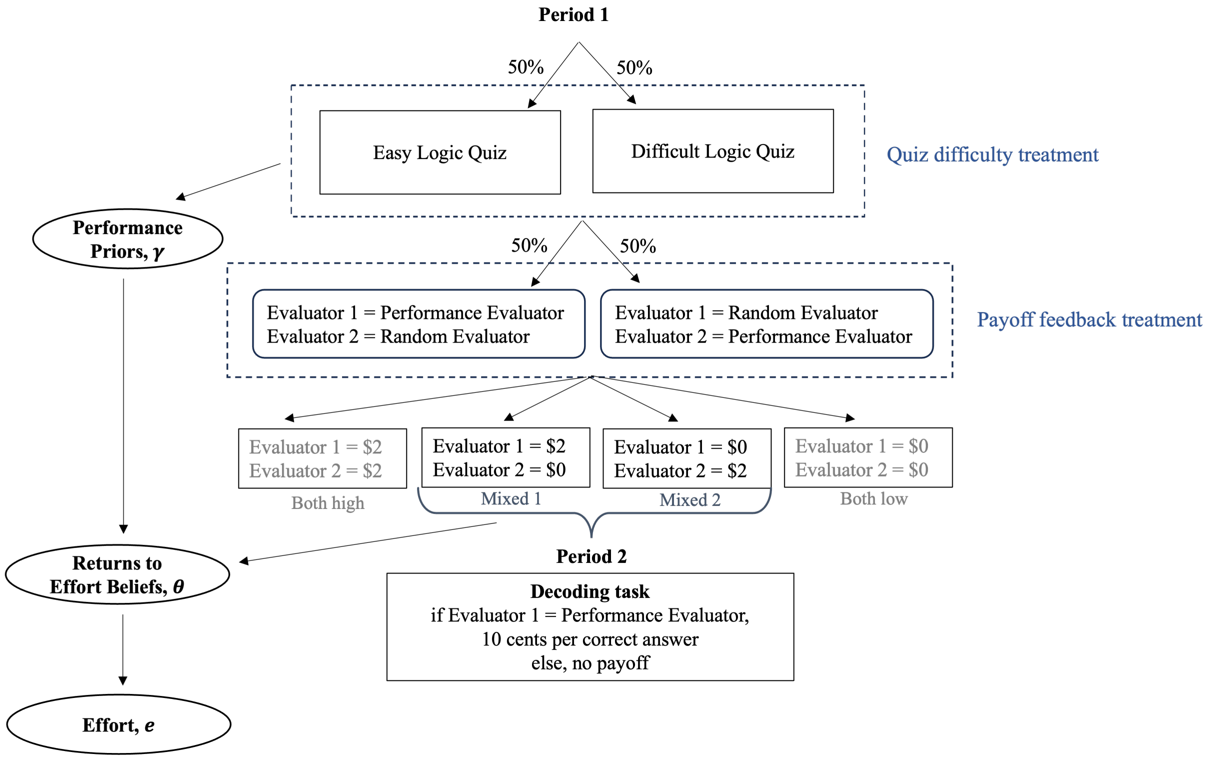


Figure 1: Experimental Design

Treatments The experiment features two between-subjects binary treatments that are assigned orthogonally and with equal probability. The purpose of the treatments for identification are explained in Section 2.2. The first treatment assigns participants to Easy and Difficult versions of the logic quiz, which we refer to as *quiz difficulty treatment*. While questions 1-7 are the same among both conditions, questions 8-12 vary in difficulty levels between the conditions. The quiz difficulty treatment provides exogenous variation in participants' beliefs about their worker type as individuals form more optimistic beliefs about their relative performance in easy compared to difficult tasks (Moore and Healy, 2008).

The second treatment randomizes the evaluator types assigned to 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 worker type, the outcome of a coin toss and the randomly assigned evaluator types. 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 their worker type. As a result, participants who receive mixed payoff feedback in period 1 cannot separately identify the contribution of their worker type and external luck, paralleling situations that Heidhues et al. (2018) and Hestermann and Le Yaouanq (2021) study. 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, we presented participants with 25 decoding tasks. Each task involves translating a 5 digit number into text based on the decoding key. We informed participants 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.[6]

We also informed participants that they are paid for their work in period 2 by the same evaluator 1 from period 1. Thus, the evaluator type 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 in the experiment) for each correctly solved decoding task.[7] 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 $\theta \stackrel{\text{def}}{=} pr(P) \in (0, 1)$ about the likelihood that evaluator 1 is the performance evaluator (P).[8] 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, and the time they spent working on them, which we use to proxy effort provision, e.

2.2 Causal Inference

In the context of this experiment, we aim to identify the impact of miscalibrated performance priors on effort provision due to misguided inference about the returns to effort from payoff feedback. Figure 2 illustrates 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. The experimental paradigm allows

[6]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).

[7]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.

[8]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.

us to identify the causal path even in the presence of unobservable confounders, because our treatments provide exogenous variation in (1) the degree of prior bias, and (2) the direction of misguided inference.

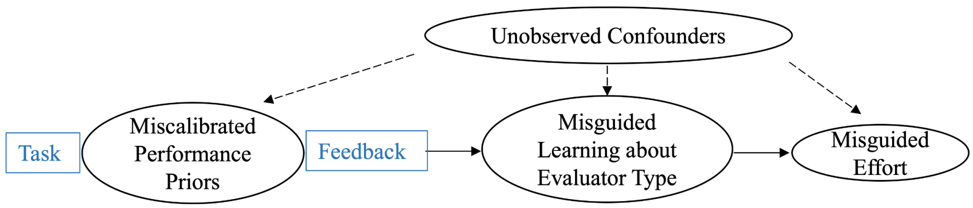


Figure 2: Representation of the Causal Mediation Chain and Treatments

Exogenous Variation in Prior Bias. Individuals are known to have miscalibrated priors about their intelligence (see Moore and Healy, 2008, for a review), which is one motivation for us to choose the logic quiz as the period 1 task. These priors are also heterogeneous across workers, however, this heterogeneity may be non-random. For example, an individual's proclivity of performing well on the experimental tasks (due to one's ability, concentration, or motivation to perform, etc.) is an unobserved individual factor that can be a confounder. If the period 1 performance influences priors (which in our data, does), the individual's performance proclivity creates a spurious correlation between priors in period 1 and on effort provision in period 2. This unobserved heterogeneity may also shape differences in belief updating, leading to a confounding path between the degree of misguided inference and performance in either period. Therefore, we create exogenous variation in the degree of prior miscalibration across subjects by randomly presenting either an Easy or Difficult version of the intelligence quiz. We keep the two versions of the logic quizzes the same for the first 7 question and vary whether the last 5 questions are easy or difficult. This allows us to obtain a measure aptitude on the logic quiz across all participants. The Easy/Difficult logic quiz difficulty treatment acts as an instrumental variable that shifts prior bias, as participants form more optimistic beliefs about their relative performance in the Easy condition compared to the Difficult condition (Moore and Healy, 2008; Möbius et al., 2022).

Exogenous Variation in Misguided Inference. One innovation of this paper is to design a feedback treatment that can act as an instrumental variable that shifts the direction of misguided inference about evaluator type without inducing an alternative causal effect on effort provision that does not operate through misguided inference. Commonly used feedback treatments in the literature either create variation in earnings and/or provide information about the individual's relative rank or intelligence. Such treatments are not likely to satisfy the exclusion restriction, and thus threaten the validity of the experimental manipulation as an instrumental variable. In particular, we worry that variations in earnings or ego-relevant information can directly impact effort provision through creating differential mood or motivational effects. Thus, we need the feedback treatment to have the following 4 features: (1) the feedback must allow (Bayesian) workers to reach different conclusions regarding the evaluator type based on their priors about their own type, and (2) the treatment must change the way workers learn about the evaluator type conditional on their priors, while (3) not induce different degrees of learning about the worker's own type and (4) not create payoff differences. First, all participants in MIXED 1 and MIXED 2 groups receive a total payoff of $2 from period 1, satisfying the last criterion. Below, we explain how the treatment also satisfies the other three criteria.

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 $s_1 = H$:

$$\begin{aligned} pr(P|s_1 = H, s_2 = L) &= \frac{pr(P)Pr(s_1 = H, s_2 = L|P)}{pr(P)Pr(s_1 = H, s_2 = L|P) + (1 - pr(P))(1 - Pr(s_1 = H, s_2 = L|P))} \\ &= \frac{0.5\gamma}{0.5\gamma + 0.5(1-\gamma)} = \gamma \end{aligned} \tag{1}$$

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

$$\begin{aligned} pr(P|s_1 = L, s_2 = H) &= \frac{pr(P)Pr(s_1 = L, s_2 = H|P)}{pr(P)Pr(s_1 = L, s_2 = H|P) + (1 - pr(P))(1 - Pr(s_1 = L, s_2 = H|P))} \\ &= \frac{0.5(1-\gamma)}{0.5(1-\gamma) + 0.5\gamma} = 1 - \gamma \end{aligned} \tag{2}$$

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, satisfying the first criterion we laid out above. In the MIXED 1 treatment, we predict θ to equal γ and in the MIXED 2 treatment, we predict θ to equal $1-\gamma$, satisfying the second criterion that the treatment must change the way workers learn about the evaluator type conditional on their priors.

Finally, we show that there is no learning about ability in the MIXED 1 and MIXED 2 groups, each receiving a high and low payoff from evaluators 1 and 2, respectively. 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,

$$
\begin{aligned}
\gamma_1(s_i = H, s_{-i} = L) &= \frac{Pr(s_i = H, s_{-i} = L|H)\gamma_0}{Pr(s_i = H, s_{-i} = L|H)\gamma_0 + (1 - Pr(s_i = H, s_{-i} = L|H))(1-\gamma_0)} \\
&= \frac{0.5\gamma_0}{0.5\gamma_0 + 0.5(1-\gamma_0)} = \gamma_0
\end{aligned}
\tag{3}
$$

This design feature is crucial to defend the plausibility of the exclusion restriction that needs to be satisfied for the manipulation to be a valid instrument, 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. The lack of learning about one's own type also motivates our choice of the task in period 1, as individuals are shown be resistant to revise their overly optimistic beliefs about their intelligence (Drobner, 2022; Drobner and Goerg, 2024; Möbius et al., 2022; Zimmermann, 2020).

2.3 Theoretical Predictions

This section provides definitions and derives theoretical predictions in the context of our experiment. 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 $\Delta\gamma \stackrel{\text{def}}{=} \gamma - \gamma^*$

We calculate the objective probability of ranking in the top half 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. Prior bias can take on any value between -1 and 1. We refer to individuals with positive (negative) prior bias as overconfident (underconfident).

The inference about the returns to effort beliefs θ is misguided to the extent that subjective performance priors γ depart from the objective performance priors 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 $\theta^* = 1 - \gamma^*$ if $s_1 = L$. We define misguided inference ($\Delta\theta$) as the difference between the participant's returns to effort belief and the objective returns to effort belief.

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

Misguided inference ranges from -1 to 1. When individuals' inferences are positively (negatively) misguided, they are more optimistic (pessimistic) about returns to effort than they should be. For Bayesian individuals, misguided inference $\Delta\theta$ is equal to $\gamma - \gamma^* = \Delta\gamma$ if $s_1 = H$ and it is equal to $\gamma^* - \gamma = -\Delta\gamma$ if $s_1 = L$. Given the relationship between prior bias ($\Delta\gamma$) and misguided inference ($\Delta\theta$), we predict

Hypothesis 1 *Overconfident individuals will be positively misguided in* MIXED 1 *and negatively misguided in* MIXED 2. *Underconfident individuals will be negatively misguided in* MIXED 1 *and positively misguided in* MIXED 2. *More generally, the differences in misguided inference between* MIXED 1 *and* MIXED 2 *increase monotonically in prior bias.*

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

$$u(e, \theta) = \theta\omega e - c(e) \tag{4}$$

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 predicted effort with misguided returns to effort beliefs θ and the predicted effort with rational returns to effort beliefs 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 predict

Hypothesis 2 *The expected utility-maximizing effort is monotonically increasing in returns to effort beliefs θ and consequently monotonically increasing in the degree of misguided inference $\Delta\theta$.*

Finally, given the relationship between prior bias ($\Delta\gamma$), misguided inference ($\Delta\theta$), and misguided effort (Δe), we can derive the following prediction pertaining to the impact of prior bias on misguided effort:

Hypothesis 3 *Overconfident individuals will provide too much effort in* MIXED 1 *and provide too little effort in* MIXED 2*. Underconfident individuals will provide too little effort in* MIXED 1 *and provide too much effort in* MIXED 2*. More generally, the difference in misguided effort between* MIXED 1 *and* MIXED 2 *groups increases monotonically in prior bias.*

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 inference and misguided effort even in the presence of unobserved confounders, such as one's potential of performing well on experimental 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. We organize the empirical findings 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 inference 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 inference. Next, in Section 3.4, we analyze the consequences of this misguided inference for participants' efforts on the decoding tasks in period 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 inference on participants' subsequent effort provision. Finally, we quantify the degree of misguided effort and analyze the causal impact of prior bias on the degree of misguided effort.[9]

3.1 Overview

We present the summary statistics of the main variables informing our analyses 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.

[9] Appendix D.1demonstrates the robustness of all table results using a sample restricted to participants who passed every comprehension check question.

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
Misguided Inference	3.7	37.6	-100	-22.2	1.8	32.5	99.9
Abs. Misg. Inference	30.4	22.4	0	11.8	26.7	46.2	100
Number of Tasks Attempted	12.8	10.0	1	3	10	25	25
Time spent working (in sec.)	216	180	1	50	174	347	968
Misguided Effort (Num. of Tasks)	0.4	3.8	-10.1	-2.3	0.2	3.3	10.1
Misguided Effort (Time Spent)	5	52.7	-140	-31	3	46	140
Abs. Misg. Effort (Num. of Tasks)	3.1	2.3	0	1.2	2.7	4.7	10.1
Abs. Misg. Effort (Time Spent)	43	31	0	17	37	65	140

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 misguided inference as the difference between the participant's returns to effort belief and the objective returns to effort belief. We calculate the objective returns to effort belief by applying Bayes' rule to the objective probability of that individual ranking in the top half. Looking at the absolute value of misguided inference, 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 participants worked on and time spent working as measures for effort put forth by the participants. On average, participants work on 12.8 tasks out of 25 available tasks and spend 215.8 seconds working. However, the distribution of effort has a large variance, as we will discuss further. We calculate misguided effort as the difference between the predicted effort with and without misguided inference about the returns to effort. The counterfactual is based on the causal estimates of the impact of misguided inference on effort provision, which we will discuss in further detail in Section 3.4. Here, we want to highlight that absolute levels of misguided effort are quite substantial: in absolute terms, the predicted effort with misguided returns to effort beliefs deviates by 3.1 decoding tasks and 42.6 seconds of work time from the predicted effort under rational returns to effort beliefs.

3.2 Miscalibrated Priors

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 grey bars and the distribution of performance priors of Difficult quiz participants with white 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 $p's < 0.001$).

Panel A: Performance Priors

Panel B: Prior Bias

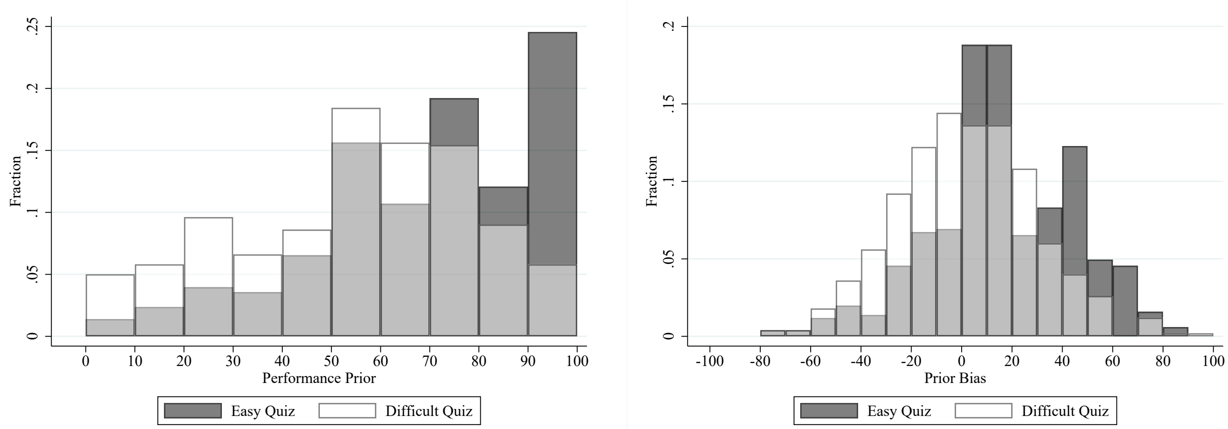


Figure 3: Manipulation check of the quiz difficulty treatment

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, $p < 0.001$).[10] 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 average prior bias between Easy

[10]Male participants are substantially more overconfident than female participants in both conditions. Appendix B explores heterogeneity with respect to gender along the causal chain from prior bias to effort provision through misguided inference.

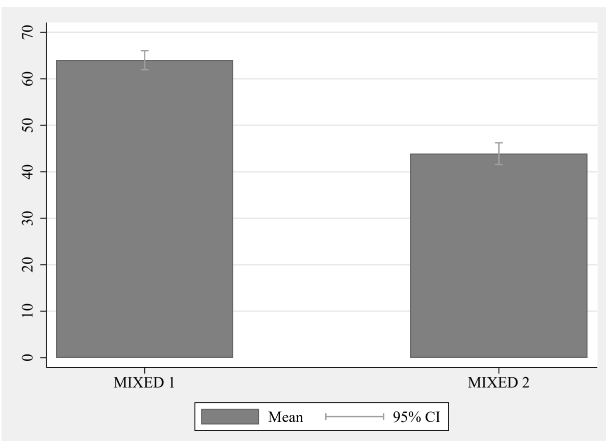
quiz and DIFFICULT quiz conditions is statistically significant (Wilcoxon rank-sum test, $p < 0.001$).[11] Next, we examine the downstream consequences of prior bias on inferences about the returns to effort.

3.3 Misguided Inference

Hypothesis 1 predicts overconfident individuals to be positively misguided in MIXED 1 and negatively misguided in MIXED 2, and underconfident individuals to be 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 inference (Panel B) across MIXED 1 and MIXED 2 groups.

Panel A: Returns to Effort Beliefs

Panel B: Misguided Inference



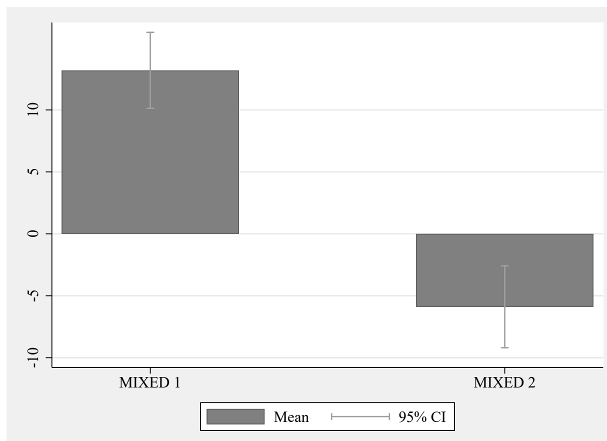


Figure 4: Returns to Effort Beliefs and Misguided Inference across MIXED 1 and MIXED 2 groups

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, $p < 0.001$). As predicted by theory, Panel B in Figure 4 shows that participants' inferences 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, $p < 0.001$). This difference of 19.1

[11] 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 ($p < 0.001$).

percentage points is statistically significant (Wilcoxon rank-sum test, $p < 0.001$). The simple model in Section 2.3 assumes Bayesian updating and therefore predicts that misguided inference in returns to effort beliefs solely arises from a bias in performance priors. However, departures from Bayesian updating (*updating bias*) may also be present. Appendix C decomposes misguided inference in misguided inference that is driven by prior bias versus deviations from Bayesian updating. The results show that misguided inference is predominantly driven by prior bias.

Next, we examine the hypothesized heterogeneous treatment effects of the payoff feedback treatment on the degree of misguided inference based on the direction of participants' prior bias. Figure 5 contrasts the means and 95% confidence intervals of misguided inference across MIXED 1/MIXED 2 conditions for overconfident ($\Delta\gamma > 0$) individuals (Panel A) and for underconfident ($\Delta\gamma \leq 0$) individuals (Panel B). As predicted by hypothesis 1, Panel A shows that overconfident individuals' inferences are positively misguided by 28.8 percentage points in MIXED 1 (Wilcoxon signed-rank test, $p < 0.001$), and negatively misguided by 20.5 percentage points in MIXED 2 (Wilcoxon signed-rank test, $p < 0.001$). The theorized opposite pattern emerges in Panel B for underconfident participants. Specifically, underconfident participants' inferences are negatively misguided by 14.5 percentage points in MIXED 1 (Wilcoxon signed-rank test, $p < 0.001$), and positively misguided by 19.1 percentage points in MIXED 2 (Wilcoxon signed-rank test, $p < 0.001$).

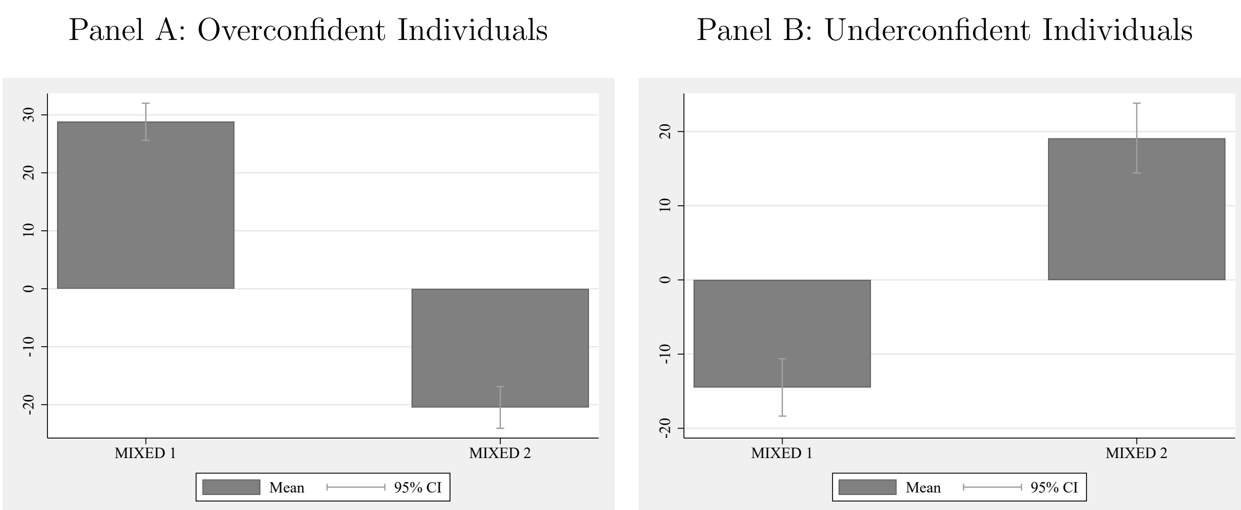


Figure 5: Misguided Inference across MIXED 1 and MIXED 2 groups, Split by the Direction of Prior Bias

More generally, hypothesis 1 predicts that the difference in the degree of misguided inference between MIXED 1 and MIXED 2 increases monotonically in prior bias. To directly test for the hypothesized heterogeneous treatment effects, we first estimate an OLS specification that regresses the degree of misguided inference 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 the degree of misguided inference between MIXED 1 and MIXED 2 groups increases in prior bias.

Table 2: Heterogeneous Treatment Effect of MIXED 1/MIXED 2 on Misguided Inference by Prior Bias

Dependent Variable:	Misguided Inference	
	(1)	(2)
MIXED 1	19.088***	19.091***
	(1.704)	(1.896)
PRIOR BIAS	-0.759***	-0.228
	(0.043)	(0.226)
MIXED 1*PRIOR BIAS	1.655***	0.771***
	(0.053)	(0.279)
Constant	-5.876***	-5.885
	(1.352)	(1.521)
Observations	1,004	1,004
Instrumental Variables	No	Yes

Notes:
(i) Results reported in column 1 are derived from OLS regressions. Results reported in column 2 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, a causal interpretation of the heterogeneity result is limited by threat of unobservable confounders that could affect (measured) performance priors and inferences about the returns to effort. Therefore, we rely on the exogenous variation in performance priors created by the quiz difficulty treatment to provide a valid test of the hypothesized heterogeneous treatment effect. In particular, we estimate a two-stage least squares regression where the 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. We report the result of this 2SLS specification for misguided inference 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 inference across the MIXED 1 and MIXED 2 groups ($p < 0.001$). This result provides direct causal evidence that the differences in misguided inference between Mixed 1 and Mixed 2 groups is monotonically increasing in prior bias.[12]

Overall, the results show that miscalibrated performance priors lead to misguided inference 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 misguided inference are moderated by prior bias. Next, we examine the consequences of misguided inference on effort provision in period 2.

3.4 Misguided Effort

Hypothesis 2 predicts that misguided inference has a positive and causal effect on effort provision. Recall from Panel B in Figure 4 of Section 3.3 that participants' inferences in the Mixed 1 group are ceteris paribus more positively misguided 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 misguided inference causally affects effort provision.

Panel A in Figure 6 shows the distributions of the work amount 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 test, $p = 0.003$).[13] On average, participants in the Mixed 1 group worked on 13.5 decoding tasks, while participants in the Mixed 2 group worked on 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 that

[12] Although we focused on the impact of *prior bias* on *misguided inference* and the impact of *misguided inference* on *effort provision* when deriving the theoretical predictions in Section 2.3, it is easy to confirm that the model would yield qualitatively the same predictions for the impact of *performance priors* on *returns to effort beliefs* and the impact of *returns to effort beliefs* on *effort provision*. The results in Appendix D.2 confirm this conjecture

[13] 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.

misguided inference causally affects effort provision.

Panel A: Number of Tasks

Panel B: Time Spent

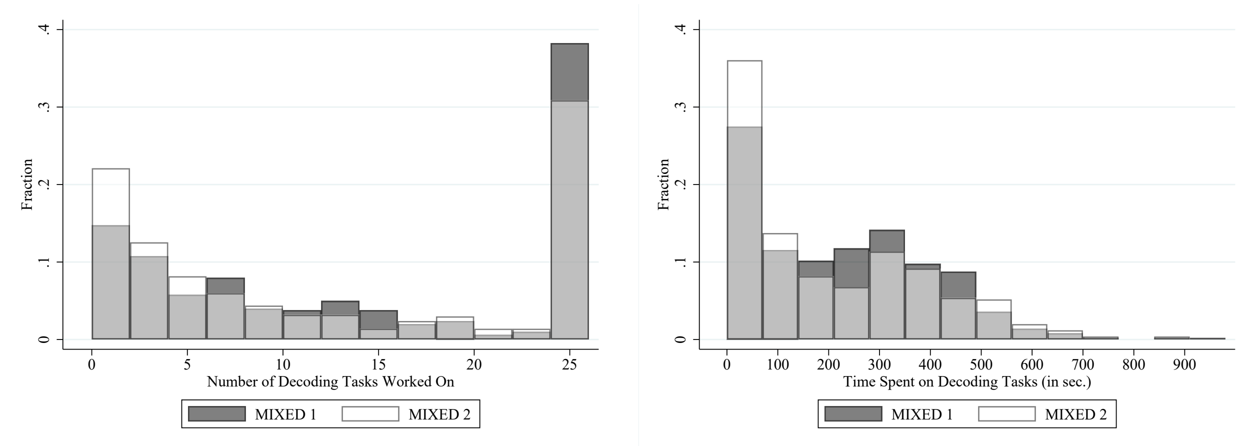


Figure 6: Effort Provision across Mixed 1 and Mixed 2 groups

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 come with higher effort costs. Next, we quantify the effect of misguided inference on effort provision, and characterize deviations from the optimal level of effort provision as a consequence of initial prior bias.

Table 3 report results from regressing the number of decoding tasks solved on the degree of misguided inference. In column 1, we first report results from an OLS specification. The significantly positive coefficient confirms that a higher degree misguided inference is associated with higher effort provision. To provide an estimate of the causal effect of misguided inference on effort provision, we rely on the exogenous variation in returns to effort beliefs created by the payoff feedback treatments and the quiz difficulty manipulation. Specifically, we estimate a 2SLS regression of the number of decoding tasks solved on the degree of misguided inference, 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 coefficient reported in column 2 shows that a 10 percentage points increase in the degree of misguided inference leads to 1.1 additionally

Table 3: The Causal Impact of Misguided Inference on Effort Provision

Dependent Variable:	Number of Tasks		Time Spent	
	(1)	(2)	(3)	(4)
Misguided Inference	0.023***	0.108***	0.394**	1.402**
	(0.008)	(0.033)	(0.155)	(0.589)
Constant	12.716***	12.405***	214.340***	210.656***
	(0.318)	(0.345)	(5.667)	(6.260)
Observations	1004	1004	1004	1004
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 misguided inference is 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, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

solved decoding tasks ($p = 0.001$).[14]

Next, we quantify the degree of misguided effort. We use the coefficient estimates for misguided inference in column 2 of Table 3 to predict the level of effort provision with and without misguided returns to effort beliefs. We call the difference between these predictions *misguided effort*. Pooling overconfident and underconfident participants, the predicted effort differs by 3.3 decoding tasks (in absolute terms) from the predicted effort with rational returns to effort beliefs based on perfectly calibrated performance priors. Recall that participants worked on 12.8 tasks on average. In this context, the extent of misguided effort is economically significant.

Figure 7 plots the means and 95% confidence intervals of misguided effort across Mixed 1/Mixed 2 conditions for overconfident ($\Delta\gamma > 0$) individuals (Panel A) and for underconfident ($\Delta\gamma \leq 0$) individuals (Panel B). As predicted by hypothesis 3, Panel A shows that, on average, compared to the predicted number of tasks they would have worked on had they held accurate priors, overconfident individuals work on 3.1 more decoding tasks in Mixed 1 (Wilcoxon signed-rank test, $p < 0.001$), and 2.2 less decoding tasks in Mixed 2 (Wilcoxon signed-rank test, $p < 0.001$). The theorized opposite pattern emerges in Panel B for underconfident participants. Specifically, underconfident participants, on average, work on 1.6 less

[14] In Appendix D.3, we reproduce the results from a robustness check that accounts for the count nature of the number of tasks by estimating Poisson regressions. The results from the IV regression suggest that a 10 percentage points increase in the degree of misguided inference leads to 0.9 additionally solved decoding tasks (p=0.002).

Panel A: Overconfident Individuals   Panel B: Underconfident Individuals

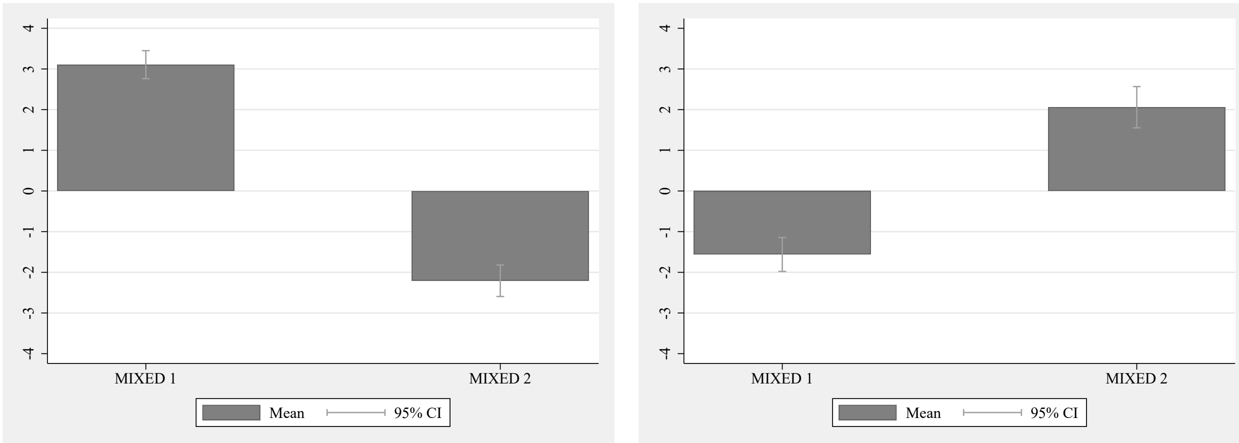


Figure 7: Misguided Effort (Number of Tasks) across MIXED 1 and MIXED 2 groups, Split by the Direction of Prior Bias

decoding tasks in MIXED 1 (Wilcoxon signed-rank test, $p < 0.001$), and 2.1 more decoding tasks in MIXED 2 (Wilcoxon signed-rank test, $p < 0.001$) compared to the number of tasks they are predicted to have worked on have they held accurate priors. The results for time spent working are analogous and depicted in Figure A4 of Appendix D.4.

More generally, hypothesis 3 predicts that the differences in the degree of misguided effort between MIXED 1 and MIXED 2 increase monotonically in prior bias. To test for the hypothesized heterogeneous treatment effects, we first present results from an OLS regression (column 1 of Table 2) that regresses our measure of misguided effort on an indicator for MIXED 1, the degree of prior bias and the interaction of the two, and then from a 2SLS regression (column 2 of Table 2) where we use the quiz difficulty treatment and the interaction between the quiz difficulty and the payoff feedback treatments as instruments for the potentially endogenous variables that are functions of performance priors. We repeat these analyses for time spent working on columns 3 and 4, respectively. We find that a 10 percentage point increase in prior bias on average leads to an increase in the misguided effort difference cross the MIXED 1 and MIXED 2 groups, corresponding to 0.83 more tasks ($p < 0.001$) and 11 more seconds of time spent working ($p < 0.001$). These results provide direct causal evidence that the differences in misguided effort between Mixed 1 and Mixed 2 groups are monotonically increasing in prior bias.

Table 4: Heterogeneous Treatment Effect of MIXED 1/MIXED 2 on Misguided Effort by Prior Bias

Dependent Variable:	Misguided Effort (Number of Tasks)		Misguided Effort (Time Spent)	
	(1)	(2)	(3)	(4)
MIXED 1	2.057***	2.058***	26.768***	26.772***
	(0.184)	(0.204)	(2.389)	(2.659)
PRIOR BIAS	-0.082***	-0.025	-1.064***	-0.320
	(0.005)	(0.024)	(0.060)	(0.317)
MIXED 1*PRIOR BIAS	0.178***	0.083***	2.321***	1.081***
	(0.006)	(0.030)	(0.074)	(0.391)
Constant	-0.633***	-0.634***	-8.240***	-8.253***
	(0.146)	(0.164)	(1.896)	(2.133)
Observations	1004	1004	1004	1004
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, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

4 Conclusion

We examine the impact of miscalibrated beliefs about one's own ability on economic behavior that is mediated through misguided inferences about the economic environment. The results show that overconfident individuals attribute poor initial labor market outcomes to low returns to effort in the economic environment, and therefore, put less effort into a subsequent real-effort task. Underconfident individuals also learn in a misguided manner and adjust their efforts accordingly, but in the opposite direction.

Extensive research in both psychology and economics demonstrates that individuals often form miscalibrated beliefs about one decision-relevant aspect and examines the direct impact of these misperceptions on economic behavior. Our findings highlight the hidden economic implications of having unrealistic expectations about one decision-relevant aspect by inducing biased judgments regarding another decision-relevant aspect.

To establish causality, our experimental design deliberately facilitates inferences about the economic environment while precluding learning about one's ability, thus eliminating any confounding effects on effort provision. However, we remain interested in how beliefs would evolve in the long-run and in settings where feedback might also contain information about ability. We hope that our study will stimulate further research aimed at developing experimental paradigms that preserve the clean causal inference property but allow beliefs to vary across multiple dimensions.

References

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

Ba, C. and A. Gindin (2023): "A multi-agent model of misspecified learning with overconfidence," *Games and Economic Behavior*, 142, 315–338.

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.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186.

Bohren, J. A. and D. N. Hauser (2021): "Learning with heterogeneous misspecified models: Characterization and robustness," *Econometrica*, 89, 3025–3077.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019): "Beliefs about gender," *American Economic Review*, 109, 739–773.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 133, 611–663.

Brownback, A. and M. A. Kuhn (2019): "Understanding outcome bias," *Games and Economic Behavior*, 117, 342–360.

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

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

Bushong, B. and T. Gagnon-Bartsch (2023): "Reference dependence and attribution bias: evidence from real-effort experiments," *American Economic Journal: Microeconomics*, 15, 271–308.

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.

Coffman, K., M. R. Collis, and L. Kulkarni (2023): “Stereotypes and belief updating,” *Journal of the European Economic Association*.

Coutts, A., L. Gerhards, and Z. Murad (2024): “What to blame? Self-serving attribution bias with multi-dimensional uncertainty,” *Economic Journal*.

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.

Drobner, C. (2022): “Motivated beliefs and anticipation of uncertainty resolution,” *American Economic Review: Insights*, 4, 89–105.

Drobner, C. and S. J. Goerg (2024): “Motivated belief updating and rationalization of information,” *Management Science*.

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

——— (2022): “By chance or by choice? Biased attribution of others’ outcomes when social preferences matter,” *Experimental Economics*, 25, 413–443.

Esponda, I. and D. Pouzo (2016): “Berk–Nash equilibrium: A framework for modeling agents with misspecified models,” *Econometrica*, 84, 1093–1130.

Exley, C. L. and J. B. Kessler (2022): “The gender gap in self-promotion,” *The Quarterly Journal of Economics*, 137, 1345–1381.

Exley, C. L. and K. Nielsen (2024): “The gender gap in confidence: Expected but not accounted for,” *American Economic Review*, 114, 851–885.

Frick, M., R. Iijima, and Y. Ishii (2020): “Misinterpreting others and the fragility of social learning,” *Econometrica*, 88, 2281–2328.

——— (2022): “Dispersed behavior and perceptions in assortative societies,” *American Economic Review*, 112, 3063–3105.

Fudenberg, D., G. Lanzani, and P. Strack (2021): “Limit points of endogenous misspecified learning,” *Econometrica*, 89, 1065–1098.

Fudenberg, D., G. Romanyuk, and P. Strack (2017): “Active learning with a misspecified prior,” *Theoretical Economics*, 12, 1155–1189.

Galindo, J. (2023): "Learning with Misspecified Models: The Case of Overestimation," *Working paper.*

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.

Graeber, T. (2023): "Inattentive inference," *Journal of the European Economic Association*, 21, 560–592.

Haggag, K., R. W. Patterson, N. G. Pope, and A. Feudo (2021): "Attribution bias in major decisions: Evidence from the United States Military Academy," *Journal of Public Economics*, 200, 104445.

Haggag, K., D. G. Pope, K. B. Bryant-Lees, and M. W. Bos (2019): "Attribution bias in consumer choice," *The Review of Economic Studies*, 86, 2136–2183.

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. (2022): "Motivating beliefs in a just world," *Working paper.*

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.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2022): "Managing self-confidence: Theory and experimental evidence," *Management Science*, 68, 7793–7817.

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.

Nyarko, Y. (1991): "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 55, 416–427.

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.

Zimmermann, F. (2020): "The dynamics of motivated beliefs," *American Economic Review*, 110, 337–363.

Özyilmaz, H. (2023): "Mental Models and Endogenous Learning," *Working paper*.

Appendices

A Balance Across Treatments

Columns 1 and 2 of Table A1 show the means of observable characteristics across quiz difficulty (Easy/Difficult) and payoff feedback treatments (Mixed 1/Mixed 2). Columns 3 and 4 of Table A1 show the magnitudes of the differences and the corresponding p-values. All reported differences are close to zero and p-values are at least greater than 0.34, confirming balance in observable characteristics across quiz difficulty and payoff feedback treatments.

Table A1: Confirming Balance

	(1)	(2)	(3)	(4)
	Easy (N=505)	Difficult (N=499)	Difference	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 Gender Differences

In this section, we analyze heterogeneity with respect to gender along the causal chain from prior bias to effort provision through misguided inference about the returns to effort. Previous literature has documented that men are more confident than women (Möbius et al., 2022; Exley and Nielsen, 2024), in particular in stereotypical male tasks (Bordalo et al., 2019; Coffman et al., 2023; Exley and Kessler, 2022). This gender gap in confidence has direct consequences for economic actions such as the well documented finding that women are less likely to enter competitive environments (Niederle and Vesterlund, 2007; Buser et al., 2014). Here, we explore the hidden economic consequences of a gender gap in confidence that arise solely through misguided inference about the returns to effort.

Panel A: Performance Priors

Panel B: Prior Bias

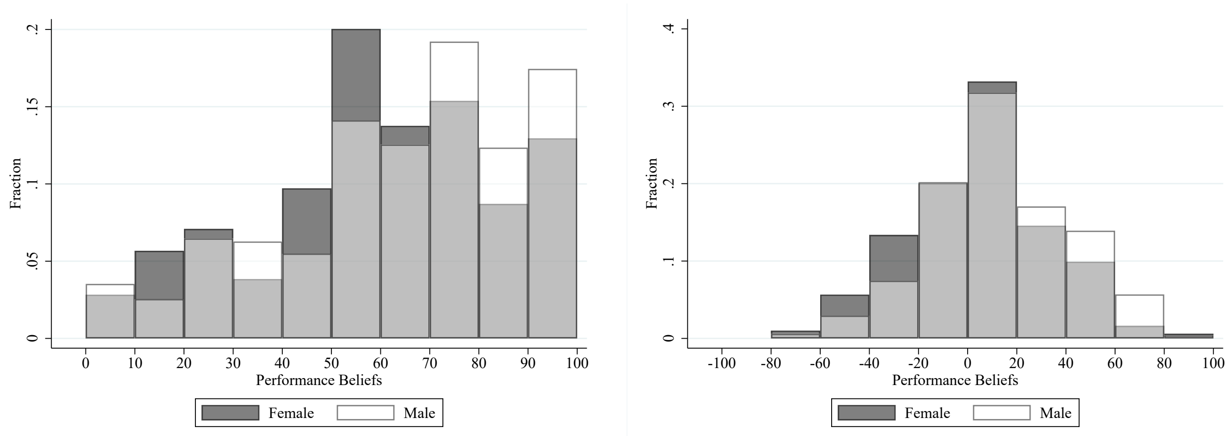
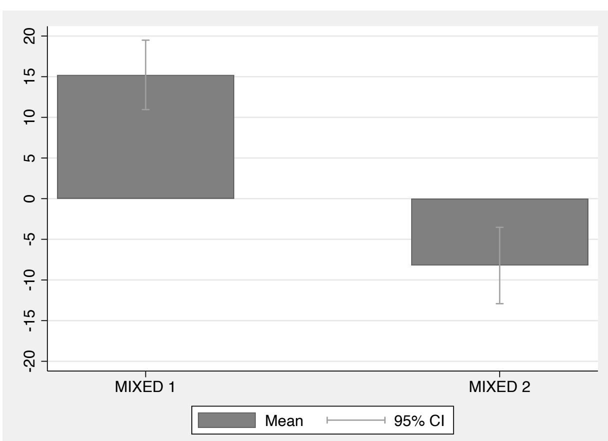


Figure A1: Gender Differences in Performance Priors and Prior Bias

First, we replicate the gender gap in confidence in our setting. Panel A of Figure 3 plots the distribution of performance priors of male participants with white bars and the distribution of performance priors of female participants with grey bars. Panel B does the same for prior bias. The distributions of performance priors and prior bias of male subjects are shifted to the right of those of female subjects (Kolmogorov–Smirnov test, both $p's < 0.001$). On average, female participants hold performance priors of 56.3%, while male participants hold performance priors of 61.8%, leading to a difference of 5.4 percentage points (Wilcoxon rank-sum test, $p < 0.001$). This gender difference in confidence increases to 7.0 percentage points when we control for participants' logic quiz scores because women tend to do better in the logic quiz than men (Wilcoxon rank-sum test, $p = 0.204$). As a result, men are substantially overconfident, with an average prior bias of 12.6 percentage points (Wilcoxon signed-rank test, $p < 0.001$) whereas women are moderately overconfident, with an average prior bias of 4.4 percentage points (Wilcoxon signed-rank test, $p < 0.001$). The 8.6 percentage points difference in prior bias between men and women is statistically significant (Wilcoxon rank-sum test, $p < 0.001$).[15]

Next, we examine the downstream consequences of this gender gap in prior bias for the extent of misguided inference about the returns to effort in period 2. Conditional on the observed gender gap in prior bias, Bayesian theory predicts that male participants would be more positively misguided in Mixed 1 and more negatively misguided in Mixed 2 than female participants. Specifically, the difference in misguided returns to effort beliefs between Mixed 1 and Mixed 2 groups is predicted to be 25.2 percentage points for men and 8.9 percentage points for women. Figure A2 depicts the actual misguided inference for Mixed 1 and Mixed 2 groups across male (Panel A) and female (Panel B) participants. Male participants are positively misguided by 15.2 percentage points in Mixed 1, and negatively misguided by 8.2 percentage points in Mixed 2 (difference of 23.4, $p < 0.001$). As predicted by theory, the difference in misguided returns to effort beliefs between Mixed 1 and Mixed 2 groups is much smaller for female participants: women are positively misguided by 11.1 percentage points in Mixed 1, and negatively misguided by 3.5 percentage points in Mixed 2 (difference of 14.6, $p < 0.001$). Overall, the difference in misguided returns to effort beliefs between Mixed 1 and Mixed 2 groups is significantly different across men and women (23.4 versus 14.6, $p = 0.054$), albeit being more muted than the difference predicted by Bayesian theory.

Panel A: Male Participants



Panel B: Female Participants

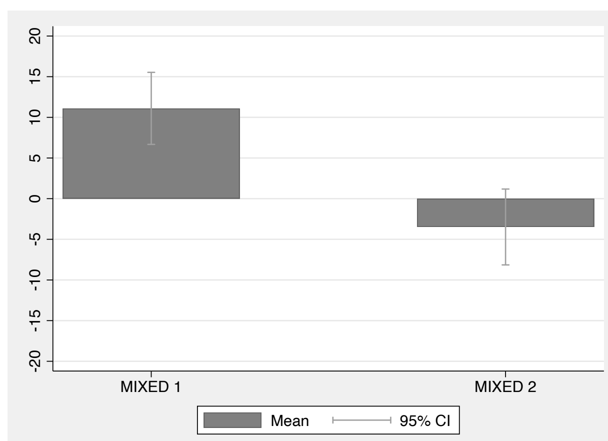


Figure A2: Misguided Inference by Gender across Mixed 1 and Mixed 2 groups

[15] Women's beliefs are more likely to be negatively biased, and the opposite is true for men. On average, men's beliefs are slightly less accurate. Among men, the average absolute departure of beliefs from reality is 25 percentage points, among women it is 23 percentage points (Wilcoxon rank-sum test, $p = 0.084$).

These misguided beliefs causally impact effort provision with similar marginal effects for men and women. Table A2 reports results from two-stage least squares regressions of the number of decoding tasks solved and time spent on the decoding tasks on two specifications. In the first specification (columns 1 and 3), we allow effort provisions to differ by gender and instrument for misguided inference with a set of interactions between dummies indicating quiz difficulty and payoff feedback treatments. In the second specification, we also allow the response to misguided inference to vary by gender by including an interaction term of misguided inference and the female dummy, and include a full set of interactions between dummies indicating gender, quiz difficulty and payoff feedback treatments as instruments. We find that women generally work more (solve on average 1.2 to 1.5 more decoding tasks and spend 33 seconds more time working) on the decoding tasks than men, regardless of their returns to effort beliefs. Misguided inference has a significantly positive effect on effort provision and the magnitudes of the coefficient estimates reported in Table 3. Importantly, the interaction term is insignificant: women's and men's effort provision in terms of number of tasks solved (column 2) and time worked (column 4) respond similarly to changes in the degree of their misguided inference.

Table A2: The Causal Impact of Misguided Inference on Effort Provision by Gender

Dependent Variable:	Number of Tasks		Time Spent	
	(1)	(2)	(3)	(4)
Misguided Inference	0.107**	0.099**	1.394**	1.319**
	(0.033)	(0.039)	(.586)	(0.667)
Female	1.155**	1.489**	33.618***	32.527***
	(0.661)	(0.703)	(11.506)	(12.711)
Misguided Inference*Female		0.023		0.292
		(0.072)		(1.270)
Constant	11.632***	11.661***	194.146***	194.410***
	(0.473)	(0.468)	(8.192)	(8.231)
Observations	1004	1004	1004	1004
Instrumental Variables	Yes	Yes	Yes	Yes

Notes:
(i) Results reported are derived from 2SLS regressions. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Using the coefficient estimates in columns 1 and 3, we predict the optimal level of effort the female and male participants would have provided if they were not misguided. Based on these predictions, we calculate the degree of misguided effort for each gender group. The gender difference in misguided returns to effort beliefs translates to a similar pattern in the degree of misguided effort provision. Men misguidedly put 21.2 seconds more work time and solve 1.6 more tasks in the MIXED 1 condition, and put 11.5 seconds less work time and solve 0.88 fewer tasks in the MIXED 2 condition (difference of 32.7 seconds and 2.5 solved tasks $p < 0.001$). Women's response is more muted: they put 15.5 seconds more work time and solve 1.2 more tasks in the MIXED 1 condition, and put 4.9 seconds less work time and solve 0.4 fewer tasks in the MIXED 2 condition (difference of 20.3 seconds and 1.6 solved tasks, $p < 0.001$). The treatment effect induced by the MIXED 1/2 manipulation is much smaller for female participants (one-sided t-test, $p = 0.027$).

C Decomposition of Misguided Inference

The Bayesian theory, presented in Section 2.3, predicts that misguided inference in returns to effort beliefs solely arises from a bias in performance priors. However, the experimental literature on belief updating has documented that people often deviate from Bayesian updating (Benjamin, 2019). Our experiment allows us to decompose how much of the misguided inference is explained by a bias in performance priors (*prior-based misguided inference*) and deviations from Bayesian updating (*updating-based misguided inference*). Figure A3 illustrates this decomposition across MIXED 1 and MIXED 2 groups. It shows that prior-based misguided inference predominantly explains total degree of misguided inference.

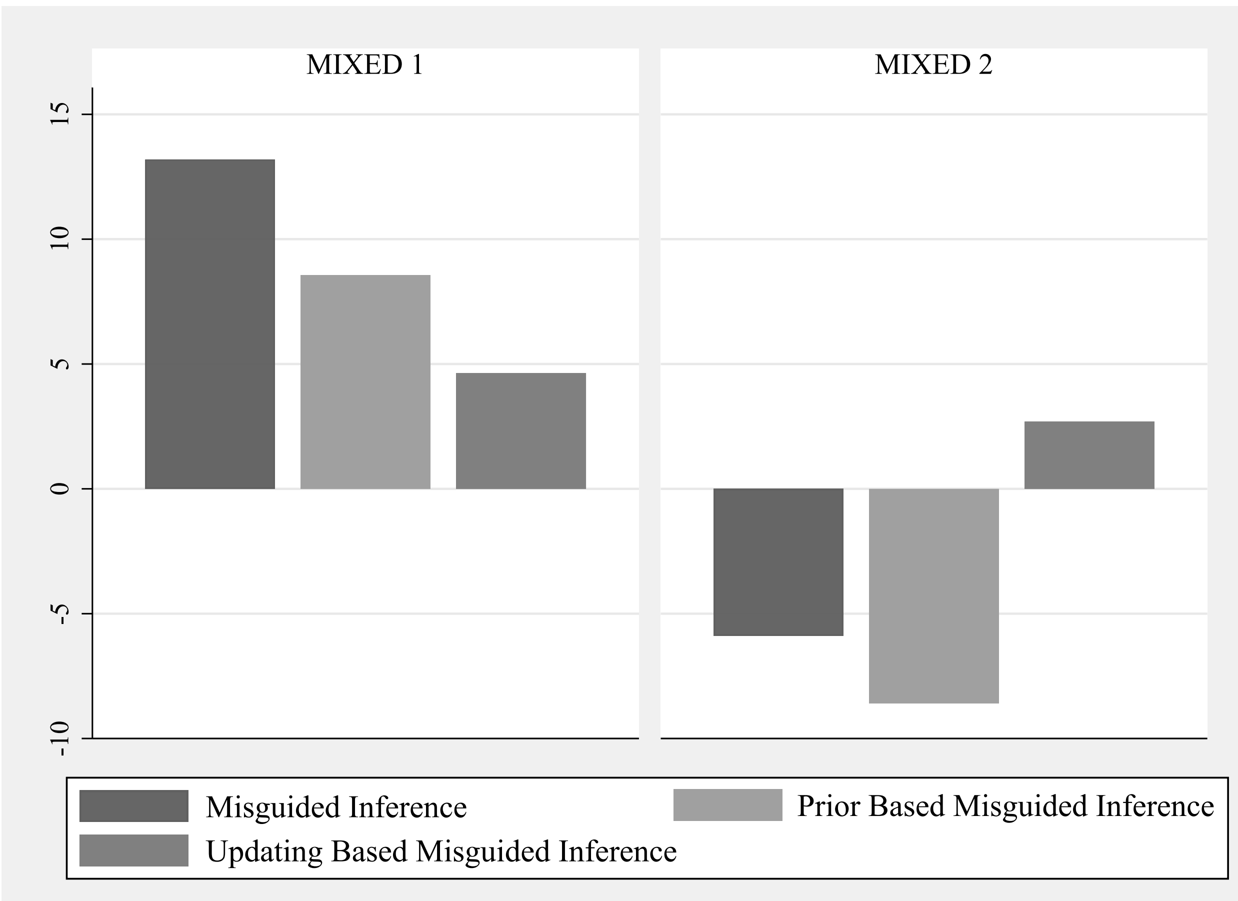


Figure A3: Decomposition of Misguided Inference

However, Figure A3 reveals that deviations from Bayesian updating also play some role in shaping the total degree of misguided inference. The updating-based misguided inference is positive across MIXED 1 and MIXED 2 groups, averaging 3.7 (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 more optimistic beliefs about the returns to effort to motivate themselves to work harder in period 2 as suggested by Bénabou and Tirole (2005) and Lobeck (2022). Second, participants may form more optimistic 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 optimistic updating bias in returns to effort beliefs indicates that the updating task provides a scope for motivated belief distortions as suggested by Coutts et al. (2024).

D Robustness Checks

D.1 Restricted Sample - Passed Comprehension Check

After explaining the source of payoffs in period 1, we asked the following two comprehension check questions. Below, we reproduce the table results of the paper, using a sample restricted to participants who successfully answered both comprehension check questions on their first attempt (82.97% of the sample).

1. *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.* ***(correct)***
   - *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.* ***(correct)***
2. *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.* ***(correct)***
   - *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.* ***(correct)***

Table A3: Heterogeneous Treatment Effect of Mixed 1/Mixed 2 on Misguided Inference by Prior Bias

Dependent Variable:	Misguided Inference	
	(1)	(2)
Mixed 1	16.280***	14.451***
	(1.816)	(2.103)
Prior Bias	-0.770***	-0.274
	(0.046)	(0.228)
Mixed 1*Prior Bias	1.713***	0.784***
	(0.056)	(0.298)
Constant	-4.268***	-3.281**
	(1.426)	(1.605)
Observations	833	833
Instrumental Variables	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. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.
(iii) Analysis uses a sample restricted to participants who successfully answered both comprehension check questions on their first attempt.

Table A4: The Causal Impact of Misguided Inference on Effort Provision

Dependent Variable:	Number of Tasks		Time Spent	
	(1)	(2)	(3)	(4)
Misguided Inference	0.024***	0.175***	0.380**	2.222**
	(0.009)	(0.057)	(0.177)	(0.967)
Constant	13.307***	12.759***	221.596***	214.929***
	(0.351)	(0.439)	(6.242)	(7.728)
Observations	833	833	833	833
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. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.
(iii) Analysis uses a sample restricted to participants who successfully answered both comprehension check questions on their first attempt.

Table A5: Heterogeneous Treatment Effect of MIXED 1/MIXED 2 on Misguided Effort by Prior Bias

Dependent Variable:	Misguided Effort (Number of Tasks)		Misguided Effort (Time Spent)	
	(1)	(2)	(3)	(4)
MIXED 1	2.856***	2.535***	36.171***	32.106***
	(0.319)	(0.369)	(4.036)	(4.672)
PRIOR BIAS	-0.135***	-0.048	-1.711***	-0.608
	(0.008)	(0.040)	(0.101)	(0.508)
MIXED 1*PRIOR BIAS	0.300***	0.137***	3.805***	1.742***
	(0.010)	(0.052)	(0.125)	(0.662)
Constant	-0.749***	-0.575**	-9.484***	-7.289**
	(0.250)	(0.281)	(3.169)	(3.566)
Observations	833	833	833	833
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. Robust standard errors are in parentheses.
(ii) Stars reflect significance in a t-test of the null hypothesis that coefficients are equal to 0, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.
(iii) Analysis uses a sample restricted to participants who successfully answered both comprehension check questions on their first attempt.

D.2 Replication of results using returns to effort beliefs

The simple model in Section 2.3 focused on the impact of *prior bias* on *misguided inference* and the impact of *misguided inference* on *effort provision.* However, it is easy to confirm that the model would yield qualitatively the same predictions for the impact of *performance priors* on *returns to effort beliefs* and the impact of *returns to effort beliefs* on *effort provision.* The results in Tables A6 and A7 confirm this conjecture.

Table A6: Heterogeneous Treatment Effect of Mixed 1/Mixed 2 on Returns to Effort Beliefs by Performance Priors

Dependent Variable:	Misguided Inference	
	(1)	(2)
Mixed 1	20.061***	20.026***
	(1.433)	(1.432)
Performance Prior	-0.312***	-0.292*
	(0.050)	(0.155)
Mixed 1*Performance Prior	0.808***	0.892***
	(0.064)	(0.190)
Constant	43.809***	43.814***
	(1.139)	(1.135)
Observations	1,004	1,004
Instrumental Variables	No	Yes

Notes:
(i) Results reported in column 1 are derived from OLS regressions. Results reported in column 2 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, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A7: The Causal Impact of Returns to Effort Beliefs on Effort Provision

Dependent Variable:	Number of Tasks		Time Spent	
	(1)	(2)	(3)	(4)
Misguided Inference	0.049**	0.177***	0.794***	3.221***
	(0.021)	(0.067)	(0.300)	(1.039)
Constant	9.852***	11.480***	182.268***	188.209***
	(0.799)	(0.573)	(11.340)	(9.291)
Observations	1004	1004	1004	1004
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 misguided inference is 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, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

D.3 Robustness of the Causal Impact of Misguided Inference on Effort Provision

Table A8: Robustness of the Causal Impact of Misguided Inference on Effort Provision

Dependent Variable:	Number of Tasks	
	(1)	(2)
Misguided Inference	0.002***	0.009***
	(0.001)	(0.003)
Constant	2.541***	2.461***
	(0.025)	(0.054)
Observations	1004	1004
Instrumental Variables	No	Yes

Notes:
(i) Results reported in column 1 are derived from a Poisson regression. Results reported in columns 2 are derived from am IV Poisson regression where misguided inference is 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, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

D.4 Misguided Effort: Time Spent

Figure A4 plots the means and 95% confidence intervals of the misguided time spent working across Mixed 1/Mixed 2 conditions for overconfident ($\Delta\gamma > 0$) individuals (Panel A) and for underconfident ($\Delta\gamma \leq 0$) individuals (Panel B). As predicted by hypothesis 3, Panel A shows that, on average, compared to the time they are predicted to have spent working had they held accurate priors, overconfident individuals spend 40.4 seconds more time on the decoding tasks in Mixed 1 (Wilcoxon signed-rank test, $p < 0.001$), and 28.7 seconds less time on the decoding tasks in Mixed 2 (Wilcoxon signed-rank test, $p < 0.001$). The theorized opposite pattern emerges in Panel B for underconfident participants. Specifically, underconfident participants, on average, spend 20.3 seconds less time working on the decoding tasks in Mixed 1 (Wilcoxon signed-rank test, $p < 0.001$), and 26.8 seconds more time working on the decoding tasks in Mixed 2 (Wilcoxon signed-rank test, $p < 0.001$) compared to the time they are predicted to spent had they held accurate priors.

Panel A: Overconfident Individuals

Panel B: Underconfident Individuals

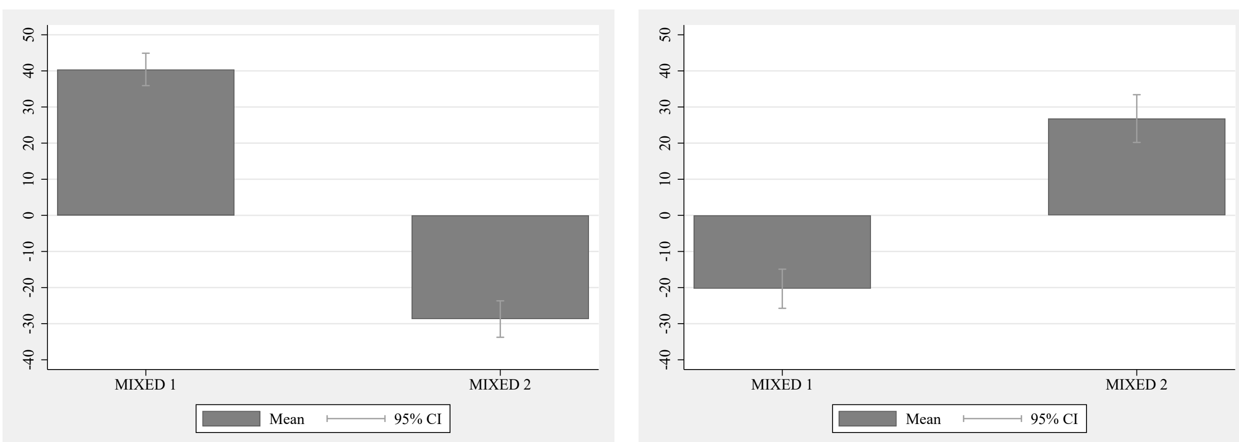


Figure A4: Misguided Work Time across Mixed 1 and Mixed 2 groups, Split by the Direction of Prior Bias

E 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.

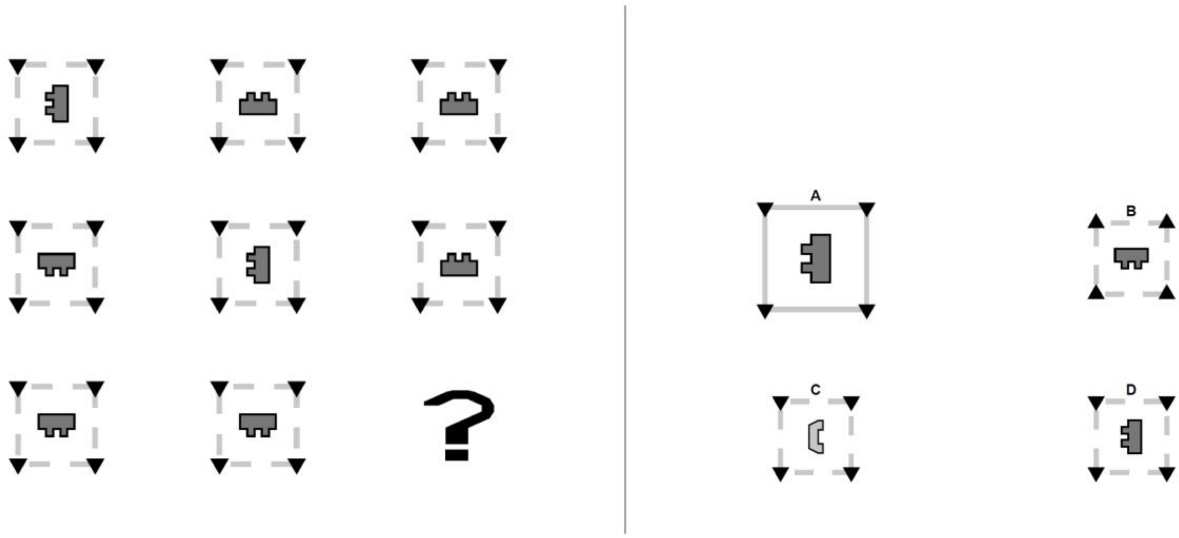
Part A

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.

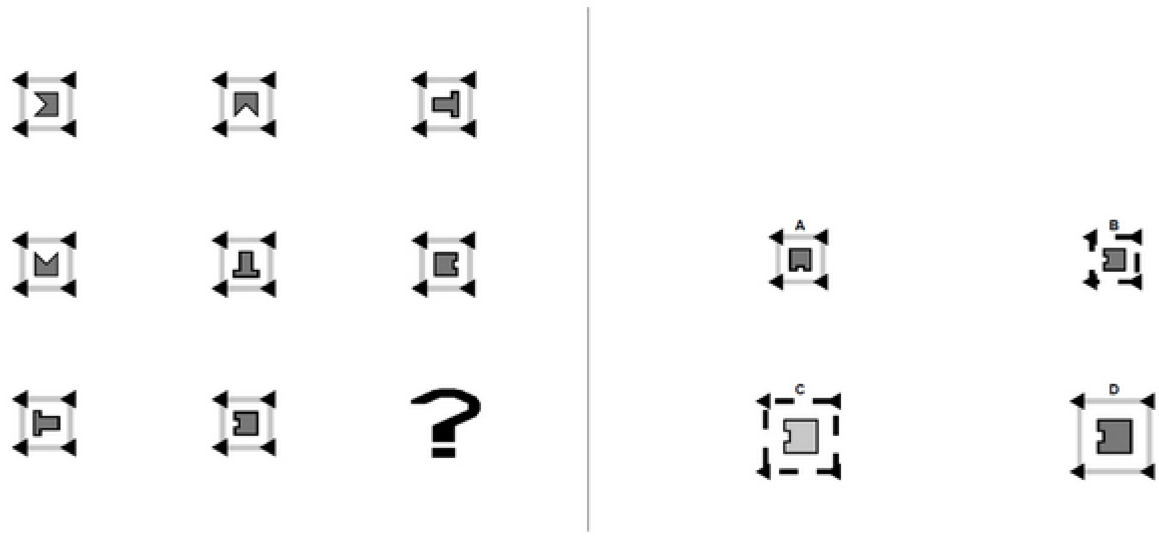
On the next page, we present a sample question of the logic quiz. The questions ask you to select the most logical option (A, B, C or D) that belongs to the ? in the pattern you see.

Sample Question



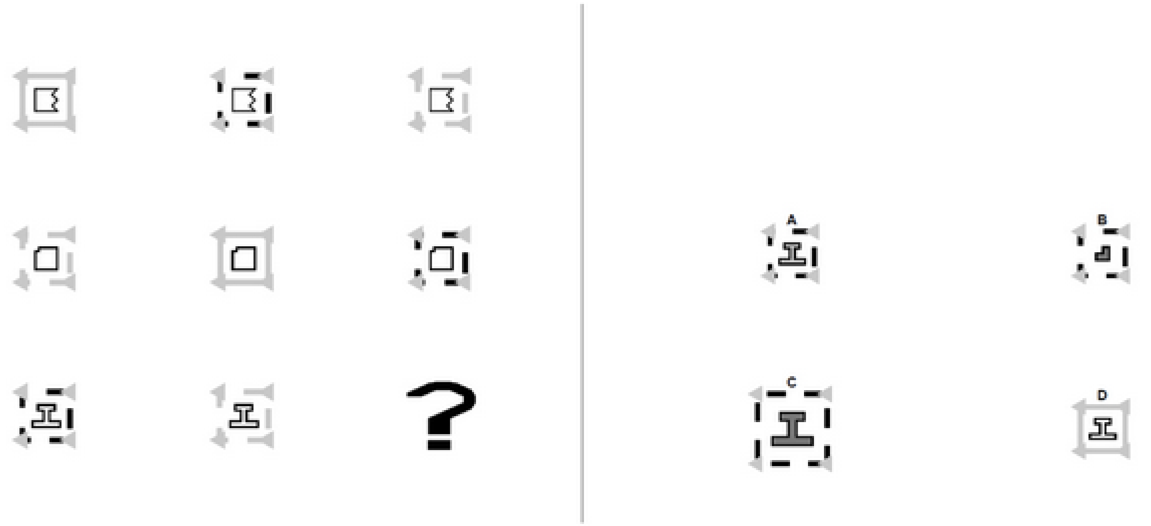
The correct answer to the sample question is Option D. You will see questions of varying difficulty. If a question feels hard to solve, it is probably a difficult question for everyone. Do your best at all times! The better you perform, the higher your payoff is expected to be. Please click the arrow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

Question (1/12)



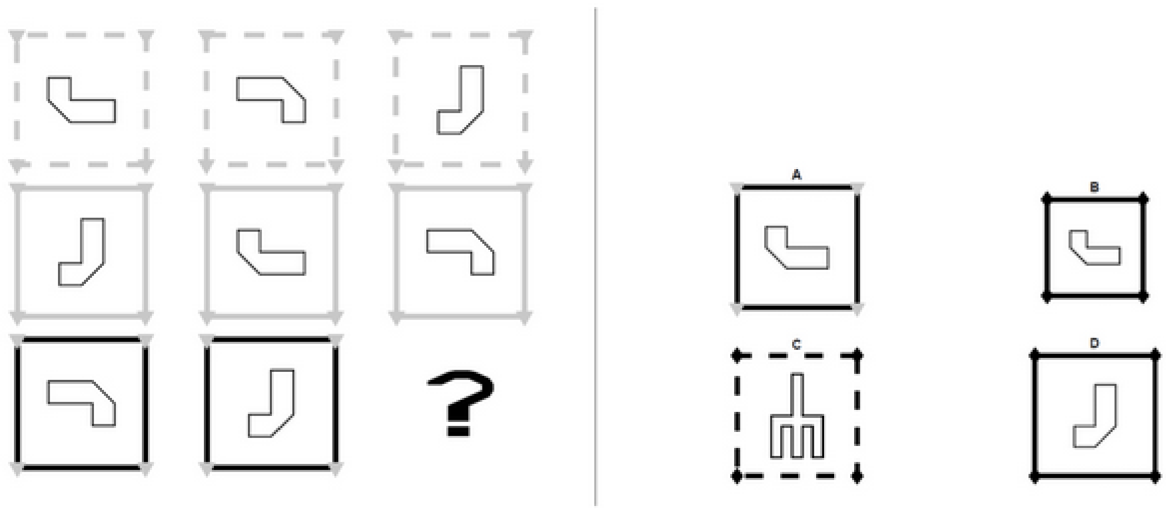
A	B	C	D
○	○	○	○

Question (2/12)



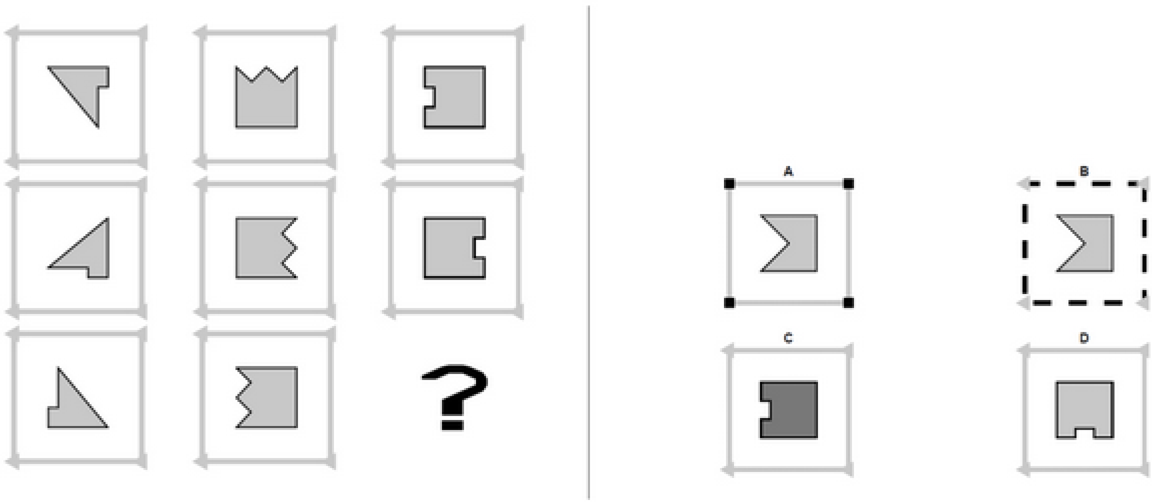
A	B	C	D
○	○	○	○

Question (3/12)



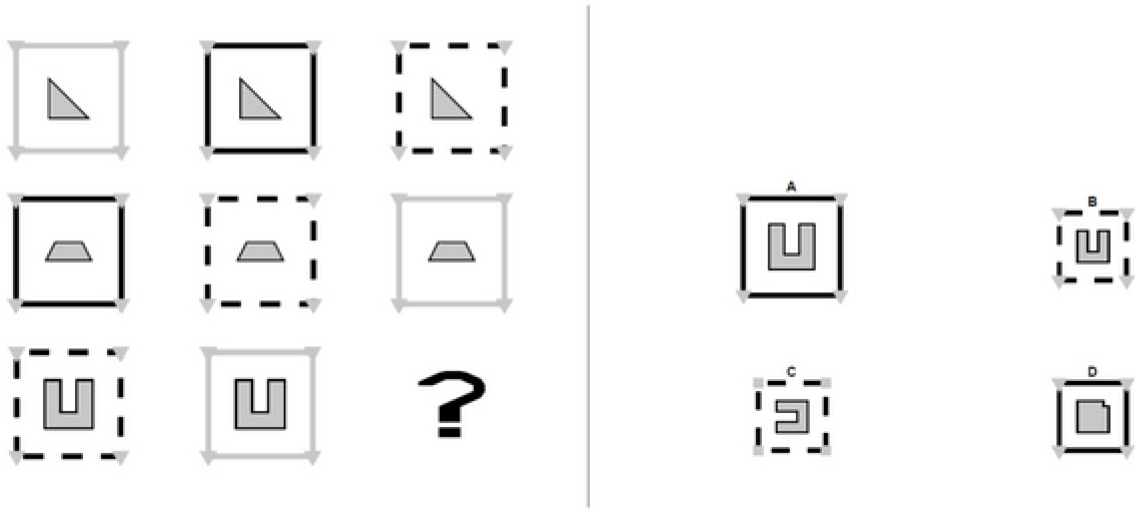
A	B	C	D
○	○	○	○

Question (4/12)



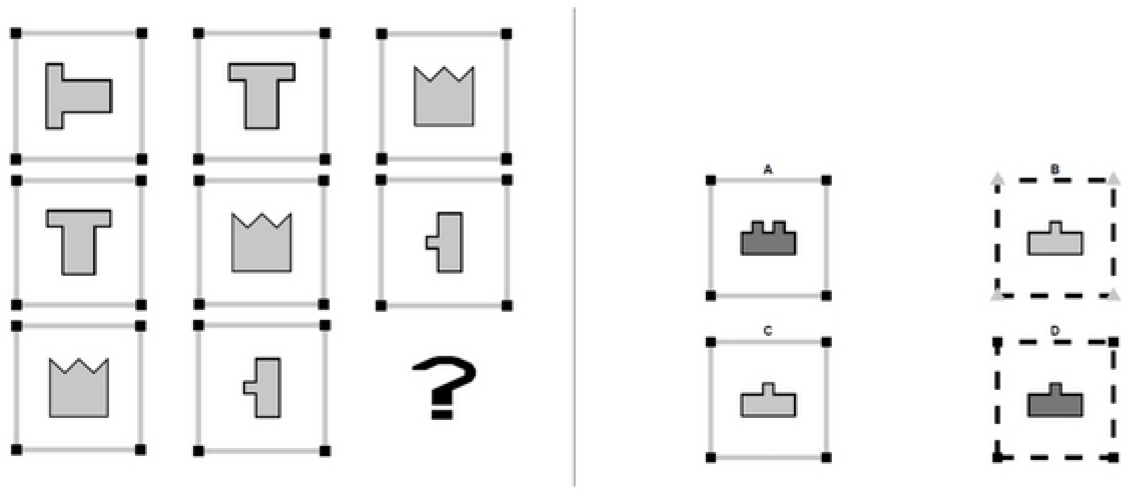
A	B	C	D
○	○	○	○

Question (5/12)



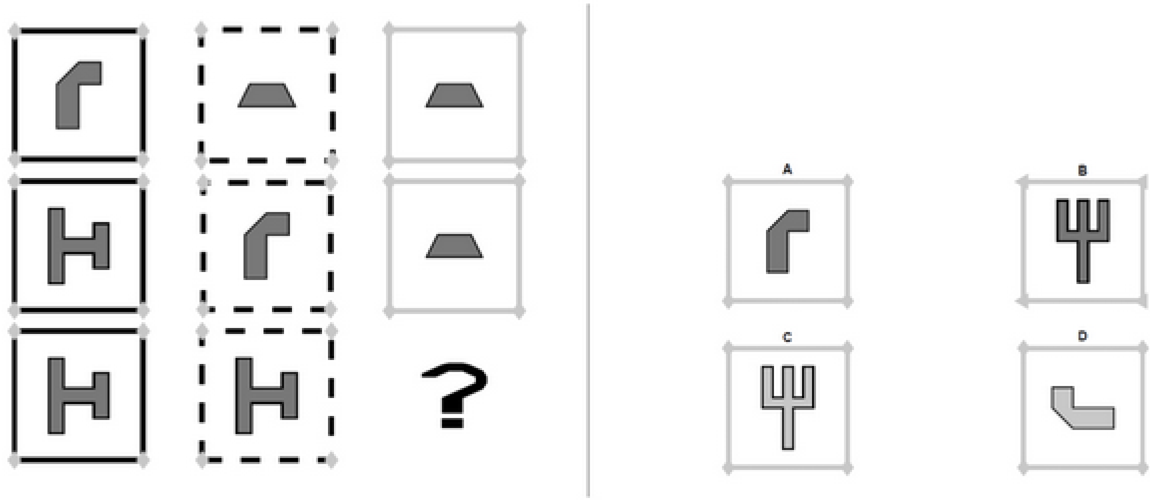
A	B	C	D
○	○	○	○

Question (6/12)



A	B	C	D
○	○	○	○

Question (7/12)

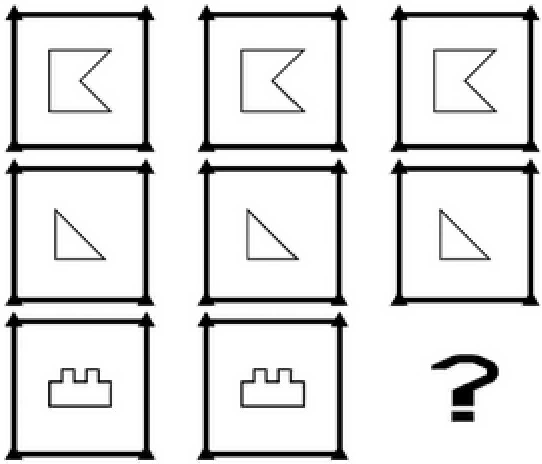


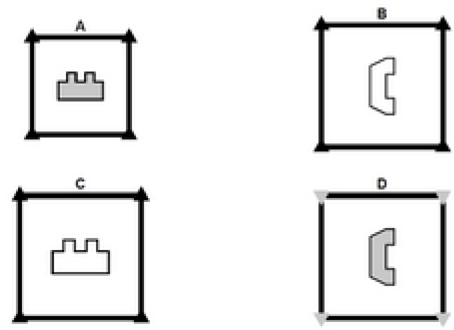
A	B	C	D
◯	◯	◯	◯

QUESTIONS 8-12 VARY BETWEEN QUIZ DIFFICULTY TREATMENTS

Question (8/12)

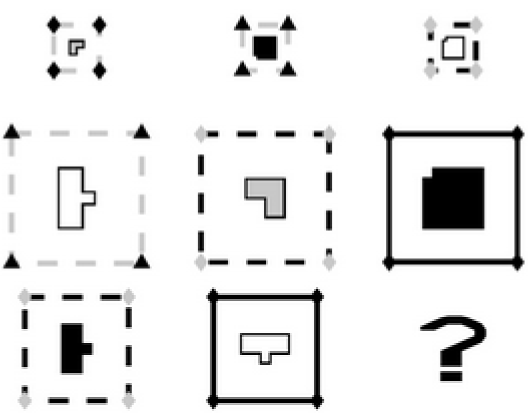
EASY

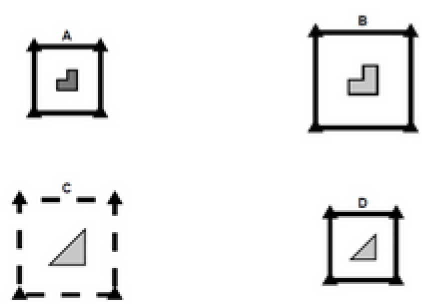




A	B	C	D
○	○	○	○

DIFFICULT

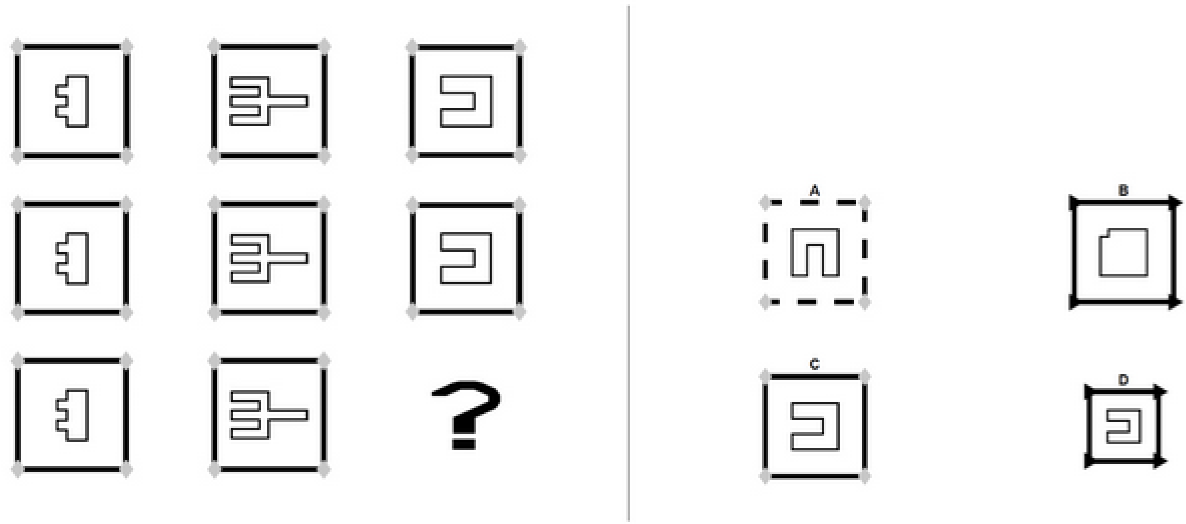




A	B	C	D
○	○	○	○

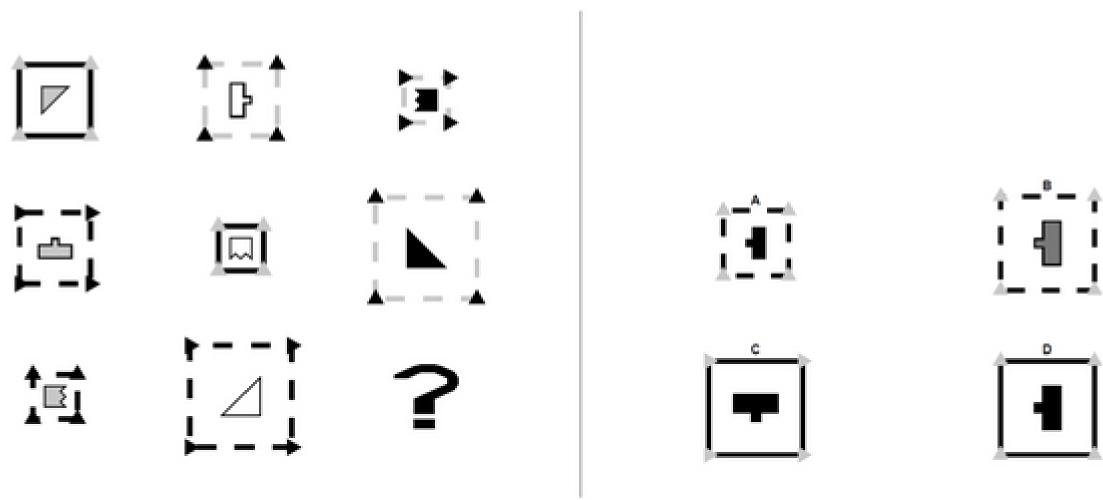
Question (9/12)

EASY



A	B	C	D
◯	◯	◯	◯

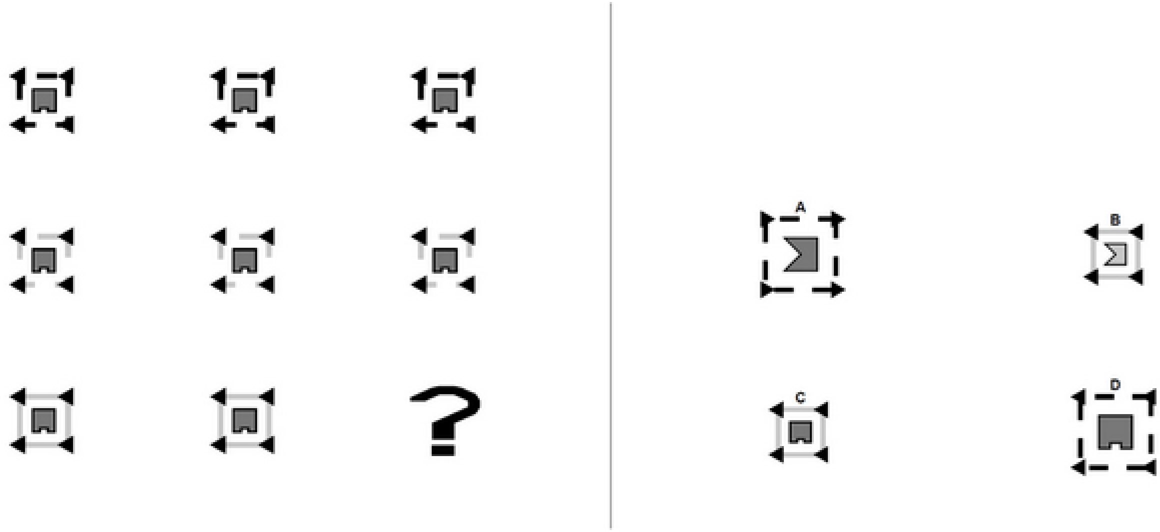
DIFFICULT



A	B	C	D
◯	◯	◯	◯

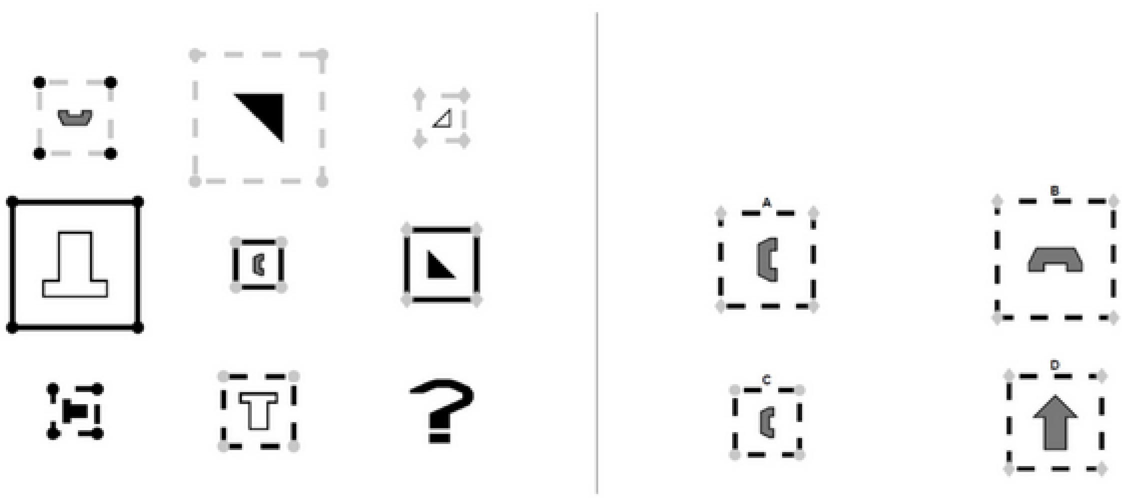
Question (10/12)

EASY



A	B	C	D
◯	◯	◯	◯

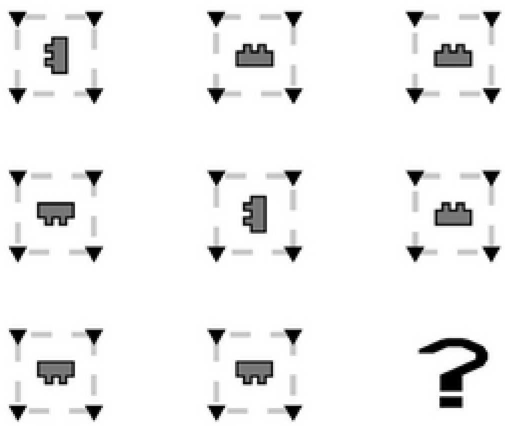
DIFFICULT

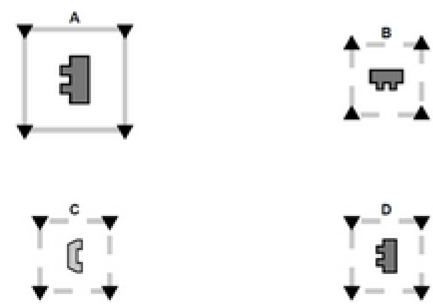


A	B	C	D
◯	◯	◯	◯

Question (11/12)

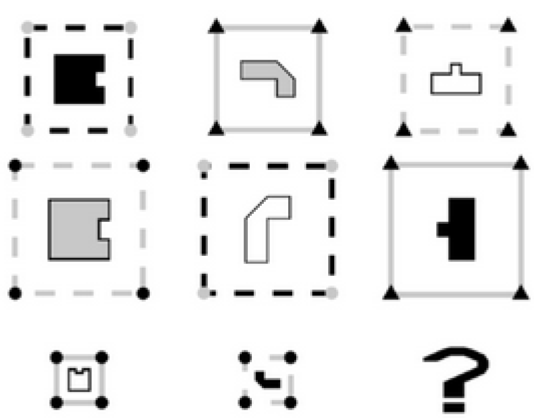
EASY

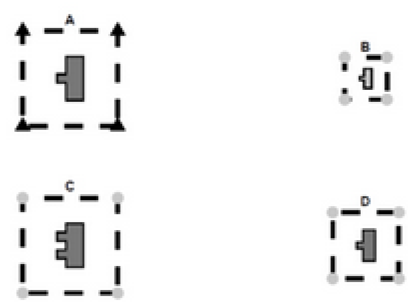




A	B	C	D
◯	◯	◯	◯

DIFFICULT

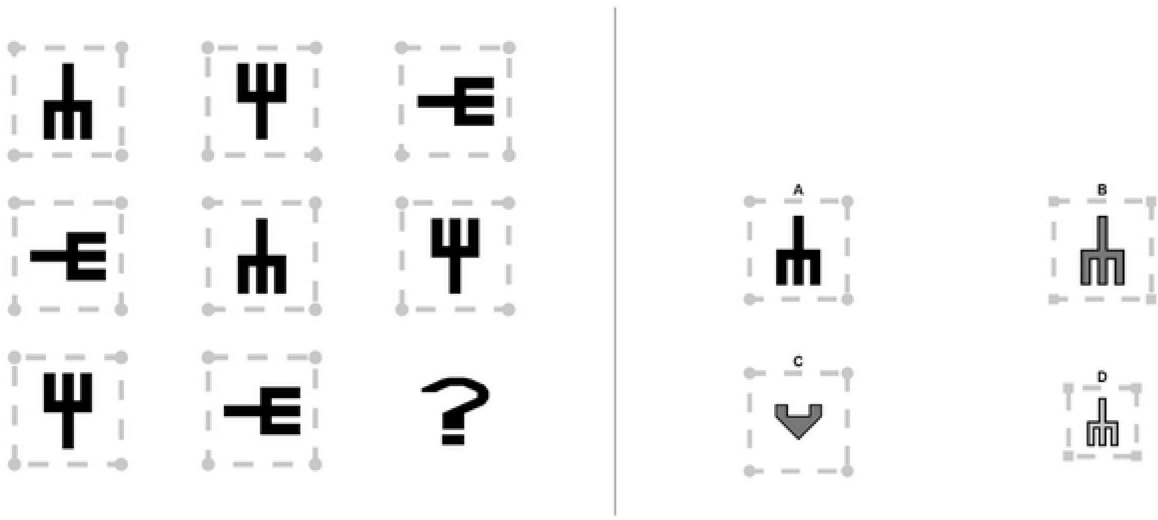




A	B	C	D
◯	◯	◯	◯

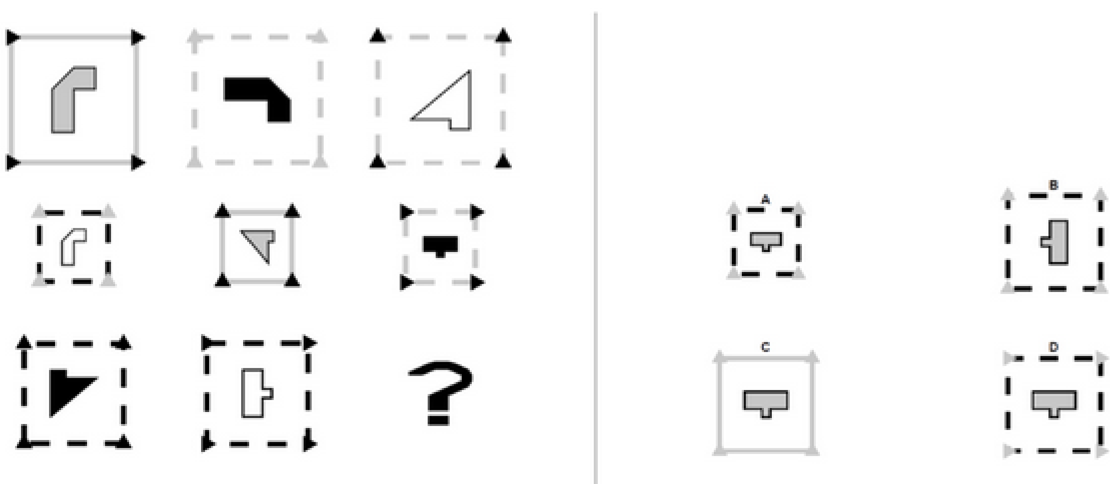
Question (12/12)

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.