When Biased Beliefs Lead to Optimal Action: An Experimental Study*

Marco Castillo[†] Sora Youn[‡]

October 20, 2023

Link to Most Current Version

Abstract

Do biased beliefs always lead to sub-optimal actions in equilibrium? Heidhues et al. (2018) demonstrate that optimal action can be achieved with misspecified beliefs when output depends not on each of the inputs independently but solely on their aggregate. This study provides an experimental test of this proposition. Supporting the theory, Experiment A highlights the exacerbated inefficiency that arises when decision-makers allocate tasks to individuals separately, guided by their potentially incorrect beliefs about the relative productivity of each person. However, this harm can be mitigated when decision-makers allocate tasks to a group of individuals, focusing solely on the average productivity of the group. Experiment B further establishes a causal link by introducing exogenous belief biases. This study holds significant implications for how to address the negative impacts of belief biases, especially when belief biases are challenging to rectify.

Keywords: Misspecified beliefs, Mental models, Learning, Lab experiments
JEL Codes: D03, D83, D91

---

*We thank to seminar and conference participants at Texas A&M University; 2022 North-American Economic Science Association Conference (Santa Barbara, CA); Southern Economic Association (SEA) 92nd Annual Meeting (Fort Lauderdale, FL), Western Economic Association International 17th International Conference (Virtual); 2023 Asia-Pacific ESA Economic Science Association Conference (Seoul, Republic of Korea); SPUDM 2023 Warm up talks (Virtual), for their helpful comments and feedback. This study is approved by the IRB at Texas A&M University (IRB2021-0481D). All errors are our own.

[†]Department of Economics, Texas A&M University. marco.castillo@tamu.edu

[‡]Department of Economics, Texas A&M University. sorayoun@tamu.edu

1 Introduction

Biased beliefs pervade our cognition, ranging from beliefs about oneself, others, and society as a whole. Studies have documented persistent biases in self-perception, which may be either overly positive or overly negative. On the one hand, individuals often maintain an excessively optimistic view of themselves (for a review, Santos-Pinto and de la Rosa, 2020). On the other hand, some individuals doubt their intellect, skills, or accomplishments despite external evidence of their competence (e.g., Clance and Imes, 1978). A substantial body of literature provides evidence that erroneous self-beliefs have significant impacts on various domains, including corporate decisions (Malmendier and Tate, 2005, 2008; Koellinger, Minniti, and Schade, 2007), education choices (Kinsler and Pavan, 2021; Stinebrickner and Stinebrickner, 2014), and labor markets (Exley and Nielsen, 2022; Hoffman and Burks, 2020; Mueller, Spinnewijn, and Topa, 2021; Niederle and Vesterlund, 2007).

Biased beliefs about others also have far-reaching implications (for a review, Bursztyn and Yang, 2022). Stereotypes, whether based on race, gender, age, or other characteristics, are associated with automatic and unconscious responses (Bodenhausen, 1990; Hilton and von Hippel, 1996; Macrae and Bodenhausen, 2000). The literature has demonstrated their influences on various aspects of society, including equal opportunities, social tensions, labor markets, and the criminal justice system (Ayres, Banaji, and Jolls, 2015; Edelman and Luca, 2014; Fiske, 2002; Lang and Spitzer, 2020). Furthermore, it is common for individuals to hold misunderstandings regarding the behaviors, beliefs, and preferences of others. These misperceptions play a fundamental role in shaping perceived social norms and influence social behavior (Bursztyn, Egorov, Haaland, Rao, and Roth, 2023; Bursztyn, González, and Yanagizawa-Drott, 2020; Gagnon-Barstch and Bushong, 2023).

People attempt to make sense of the world based on their preconceptions about oneself and others. Thus, biased beliefs about oneself and others ultimately distort our worldview. When observing a phenomenon, we may wrongly attribute it to ourselves, others, or luck depending on our preconceived perceptions (Coutts, Gerhards, and Murad, 2020; Heider, 1944; Hestermann and Le Yaouanq, 2021; Hewstone, 1990).

This extensive body of literature has posed a crucial question: Why do individuals fail to rectify their biased beliefs, even when doing so could mitigate the economic costs involved? Standard neoclassical economics suggests that rational individuals continuously adjust and revise their beliefs until they align with the correct ones as they systematically acquire, process, and integrate new information. Handel and Schwartzstein (2018) categorizes potential explanations into two categories: frictions and mental gaps.

The frictions perspective suggests that individuals may either lack access to specific information or, if accessible, rationally choose to disregard it due to the belief that the costs associated with processing the information outweigh its perceived value (Sims, 2003). In such cases, a potential solution lies in enhancing information provision and improving its accessibility (Haaland, Roth, and Wohlfart, 2020; Thaler, Sunstein, and Balz, 2012).

Conversely, when mental gaps are the root causes, reducing information friction would not effectively address the issue. Mental models refer to cognitive frameworks through which individuals comprehend and interpret the world. If these mental models are misspecified, it can lead to distortions in information-gathering, attention, and processing. For example, Hanna, Mullainathan, and Schwartzstein (2014) find that Indonesian seaweed farmers persistently failed to optimize an input dimension. Despite the availability of data, the farmers did not pay attention to that specific dimension because they failed to notice its importance.

While previous research has primarily focused on documenting these biases, examining their economic consequences, and exploring why they occur, there is limited understanding of *when* biased beliefs are harmful and *when* they are not. In this paper, we aim to tackle this question

This question holds significant importance in several ways. Firstly, it is widely acknowledged that misperceptions are pervasive, and some of them are remarkably resistant. As illustrated in the seaweed farmers' example, people often fail to recognize their mistakes, missing opportunities to learn (Enke, 2020; Enke and Zimmermann, 2019; Esponda and Pouzo, 2016; Esponda, Vespa, and Yuksel, 2022; Fudenberg, Lanzani, and Strack, 2021; Gagnon-Bartsch, Rabin, and Schwartzstein, 2021). Moreover, people are frequently motivated to maintain their existing and preferred beliefs by ignoring or misinterpreting provided information (Epley and Gilovich, 2016; Golman, Hagmann, and Loewenstein, 2017; Rabin and Schrag, 1999; Zimmermann, 2020). However, little is known about how to counteract belief biases, especially when they are persistent and hard to rectify.

Therefore, our objective is to identify conditions that ensure economically desirable outcomes while acknowledging the existence of belief biases. Understanding the conditions enables us to design institutions, including incentive structures, information systems, and more comprehensive organizational frameworks, that can safeguard against the economic harm caused by biased beliefs. To our knowledge, there are only a few papers that examine how to counteract biased beliefs, not by correcting biases, but by changing institutions. Recalde and Vesterlund (2022) advocate for initiatives to reform institutions in order to address the gender gap in negotiations. Similarly, Enke, Graeber, and Oprea (2023) explores various market institutions that facilitate the reduction of individual level cognitive biases at the aggregate level.

As one of the first steps towards this goal, this paper focuses on a task allocation problem within the context of teamwork. For efficient teamwork, it is essential to have a correct understanding of each team member's strengths. Consider a scenario where a manager runs a supermarket with two team members: Teammate 1 (TM1) and Teammate 2 (TM2). Assume that TM1 excels in customer communication, while TM2 is particularly adept at managing the store's inventory. If the manager knows their productivity for each task, the manager would strategically assign tasks to enhance the supermarket's profitability. For instance, customer-facing responsibilities can be delegated to TM1, and TM2 can take charge of handling products and inventory. However, the challenge arises because individual productivity is often not observable. In numerous instances, the manager lacks the direct means to observe and gauge the specific proficiencies of each team member. Consequently, task allocation relies heavily on the manager's subjective beliefs regarding the productivity of team members, which may be biased.

Moreover, assessing the productivity of team members is difficult, and can be misguided. Typically, the manager has access to data on the store's total profit, but distinguishing the extent to which it is attributable to customer satisfaction versus inventory management, or an exogenous economic shock is not straightforward. This lack of clarity regarding the specific contributions of team members can hinder the manager from learning the true productivity of the team members. When observing the team's profit, the manager would attribute the profit to team members based on her initial subjective beliefs about them. Therefore, if her initial beliefs are biased, the manager may misattribute, which thereby would perpetuate inefficient task assignments.

Heidhues, Kőszegi, and Strack (2018) characterize a condition where the manager can achieve efficient task allocation even when her beliefs about the team members are incorrect. Instead of assigning one of TM1 and TM2 to take full charge of customer service and the other to manage inventory, suppose the manager assigns both tasks to TM1 and TM2. In this scenario, the manager does not need to identify each team member's productivity to assign a task to a team member who can perform the best. The manager is only required to learn how well the two team members perform their jobs on average. Since it is possible to hold correct beliefs about the average productivity of the two while holding incorrect beliefs about TM1 and TM2 separately, biased beliefs are less consequential. For example, assume the manager has incorrect beliefs that TM1 excels in customer service while TM2 performs poorly, when in reality, the situation is reversed. These inaccuracies regarding each team member can offset each other when assessing the average productivity, without negatively impacting the learning of their average productivity in customer service. In summary, Heidhues et al. (2018) demonstrates if team output depends not on each team member's productivity independently but solely on the average productivity of the team members, the manager can allocate tasks optimally and achieve efficient teamwork.

We design a series of lab experiments to test this prediction. Three key features of the experiment are worth highlighting. First, it is an individual decision-making problem. As in the manager example above, every subject's main task is to allocate tasks to TM1 and TM2. To eliminate strategic concerns and interactions between team members, we pre-determine every subject's productivity (Part 1). In Part 1, each subject is assigned a productivity parameter based on their relative performance in a trivial real-effort slider task (Chen and Schildberg-Hörisch, 2019). Subjects are not provided with information about their relative performance or the absolute performance of the task. Afterward, subjects proceed to Part 2 where they state how they would allocate tasks to TM1 and TM2.

The second feature relates to team assignments. In Experiment A, each subject is assigned the role of TM1 and paired with TM2 who is a randomly selected participant. The better-than-average effect is a well-established phenomenon in which individuals tend to mistakenly believe they perform better than others, particularly on easy tasks (Moore and Healy, 2008; Zell, Strickhouser, Sedikides, and Alicke, 2020). Based on the literature, we hypothesize that subjects whose performance is below average tend to exhibit more biased beliefs about TM1, who is themselves. The choice of the trivial real-effort slider task in Part 1 is intended to accentuate the better-than-average effect.

Exploiting these naturally occurring variations in belief biases is advantageous as it allows us to draw conclusions that have real-life relevance. Nonetheless, Experiment A has a limitation in that beliefs are endogenous. For instance, subjects with below-average productivity might exhibit more biased beliefs because they are generally less attentive during the experiment. Furthermore, it's plausible that subjects may prefer to allocate more tasks to TM1 (themselves) due to a desire for control and to minimize uncertainty regarding TM2's productivity (Benoît, Dubra, and Romagnoli, 2022).

Experiment B addresses this concern. In Experiment B, every subject is paired with two other randomly selected participants, TM1 and TM2. Additionally, subjects are provided with a random signal regarding TM1's productivity. The signal reflects the fact that subjects in Experiment A have an internal signal about TM1's (themselves) productivity. It indicates whether or not TM1's productivity is below average with a 75% accuracy. More importantly, it generates random variations in belief biases. Subjects who receive incorrect signals about TM1 are likely to develop more biased beliefs about TM1, irrespective of their prior beliefs. Thus, Experiment B not only enables us to strengthen the findings from Experiment A but also establishes a causal link.

Third, the experiment is a between-subjects design. Subjects are randomly assigned to one of two treatment conditions in Part 2. In one treatment condition, referred to as Individual Task Assignment (ITA), the productivities of TM1 and TM2 independently affect team output. Subjects state how to allocate 100 hypothetical projects between TM1 and TM2. The expected team output is determined by the sum of each team member's individual output, which is proportional to the number of projects assigned to them and their productivity parameter.

In the other treatment, termed Group Task Assignment (GTA), team output depends on the average productivity of TM1 and TM2. Subjects state how to allocate 100 hypothetical projects to two distinct parties: the group consisting of TM1 and TM2 and a robot player whose productivity is known. Similar to Individual Task Assignment (ITA), the expected team output is determined by the sum of the individual outputs of each party.[1]

Individual Task Assignment (ITA) and Group Task Assignment (GTA) are comparable. In both treatments, subjects divide 100 hypothetical projects between two parties while facing two unknown parameters: TM1's and TM2's productivity. In each of the 50 rounds in Part 2, subjects only observe the team's overall output, which depends on TM1 and TM2's productivity and the subject's task allocation choice. The team output is further confounded by additive random noise. Each team member's specific contribution to the team output remains unobservable.

Consistent with the theoretical prediction, we find that biased beliefs about TM1 lead to more inefficient outcomes in Individual Task Assignment (ITA) than in Group Task Assignment (GTA). In Experiment A, subjects whose productivity is below average, and consequently hold more biased beliefs, incur a greater output loss compared to the others. However, this output loss is substantially reduced to nearly 40% in Group Task Assignment. We replicate these findings in Experiment B. Receiving incorrect signals about TM1 significantly increases output loss in ITA, but this effect decreases to nearly a quarter in GTA.

Furthermore, we demonstrate that the treatment effects are driven by different learning objectives imposed by the treatments. Subjects in the Individual Task Assignment (ITA) are required to learn the productivity ratio of TM1 and TM2 to make the optimal task allocation. In contrast, Group Task Assignment (GTA) necessitates learning the average productivity of TM1 and TM2. We find that subjects in ITA have better-calibrated beliefs about the productivity ratio, and subjects in GTA have better-calibrated beliefs about the average productivity. Crucially, we show that incorrect beliefs about TM1 hinder learning about the productivity ratio between TM1 and TM2 more than they hurt the learning of the average productivity of TM1 and TM2. As a consequence, holding biased beliefs about TM1 distorts optimal task allocation in ITA, but task allocation choices are little affected in GTA.

[1] Heidhues et al. (2018) call the first treatment *Identifiability* and the second *Non-identifiability*.

Finally, we investigate the stability of these beliefs. According to the theoretical framework, subjects optimize task allocation based on their potentially biased beliefs and form expectations about team output. If this perceived expected team output aligns with the actual expected team output, determined by the true productivities of TM1 and TM2, there is no incentive for them to further revise their beliefs: the beliefs are stable. However, our experiments provide little evidence of stable beliefs, as subjects consistently encounter significant disparities between their expectations and reality.

This paper relates to the literature on learning with incorrect mental models. The primary focus of the literature is to study the role of mental models in learning but gives less attention to strategies for rectifying them (Enke and Zimmermann, 2019; Esponda et al., 2022; Gagnon-Bartsch et al., 2021; Graeber, Forthcoming; Hanna et al., 2014; Kendall and Oprea, forthcoming; **?**). While it is not their main question, the findings of Esponda et al. (2022) suggest that policies that hide information can promote optimal behavior, as it limits the negative influence of incorrect mental models. Our findings are in line with their findings. Group Task Assignment (GTA) encourages subjects to care exclusively about the average productivity of TM1 and TM2 while blocking the acquisition of information about each team member's productivity, separately. This property is the key that makes GTA protective against biased beliefs.

There has also been a growing theoretical focus on incorrect mental models, often referred to as misspecified beliefs. The literature conceptualizes a Berk-Nash equilibrium that characterizes steady-state behavior in the presence of misspecified beliefs Esponda and Pouzo (2016); Fudenberg and Lanzani (forthcoming); Gagnon-Bartsch et al. (2021). We contribute to the literature by providing empirical validation of this framework. To the best of our knowledge, there are only two experimental papers that aim to study the empirical implications of the theory. While these two papers also draw upon Heidhues et al. (2018), our paper concentrates on a prediction that they did not investigate. Another interesting prediction from Heidhues et al. (2018) is that overconfidence leads to pessimistic views about one's surroundings. In our context, subjects with overly optimistic beliefs about TM1's productivity would be expected to underestimate TM2's productivity. Marray, Krishna, and Tang (2020) provides supporting evidence for this prediction, while Goette and Kozakiewicz (2022) reports limited evidence. Although our main focus is different, we also do not find a negative relationship between beliefs about TM1 and TM2. Indeed, we do not find evidence that beliefs converge to stable states as the theory predicts. This inconsistency offers potential avenues for future research, such as modifications of the theoretical framework to better explain real-world observations.

2 Theoretical Framework

This section analyzes the decision problem faced by subjects in our experiment. The analysis is built on Proposition 5 in Heidhues et al. (2018) and is the foundation for our experimental design.

2.1 Setting

The objective of the decision-maker (DM) is to maximize the expected output μ that is a function of her action $x \in [0, 100]$ and two unknown parameters a_1, $a_2 \in \{10, 30, 50, 70, 90\}$. Let b_1, $b_2 \in \{10, 30, 50, 70, 90\}$ denote the DM's point beliefs (or best guesses) about a_1 and a_2, respectively. For instance, a_1 can represent the productivity of one team member (TM1), and a_2 can represent the productivity of the other team member (TM2). b_1 and b_2 represent the DM's beliefs about the productivity of TM1 and TM2, respectively. The variable x denotes the work allocation between the two team members. The actual, observable output y is the sum of the expected output μ and a random error $\epsilon \sim N(0, 100)$.

Heidhues et al. (2018) define *Surprise* as the difference between the actual expected output and her perceived expected output. They demonstrate that under these two conditions, the DM's beliefs are stable. First, the decision-maker (DM) maximizes the expected output conditional on b_1 and b_2. Second, the subjectively optimal action yields the (actual) expected output that coincides with her perceived expected output. In words, beliefs are stable when expectation matches reality.[2]

$$\mathit{Surprise} = \mu(x|a_1, a_2) - \mu(x|b_1, b_2)$$

Definition 1. If the DM maximizes perceived expected output, and *Surprise* is zero, her beliefs are stable:

$$\mu(x^e|a_1, a_2) = \mu(x^e|b_1, b_2) \quad \text{such that} \quad x^e = \arg\max \mu(x|b_1, b_2)$$

More importantly, they show that the optimal action, given the stable but biased beliefs, maximizes the actual expected output if μ depends only on a summary statistic of a_1 and a_2 and not independently on the two parameters. Our treatment conditions vary accordingly.

- **Individual Task Assignment (ITA)**. μ depends independently on a_1 and a_2. The DM must acquire knowledge of the ratio between a_1 and a_2.

[2] Heidhues et al. (2018) also show that a stable belief and the corresponding optimal action constitutes a Berk-Nash equilibrium.

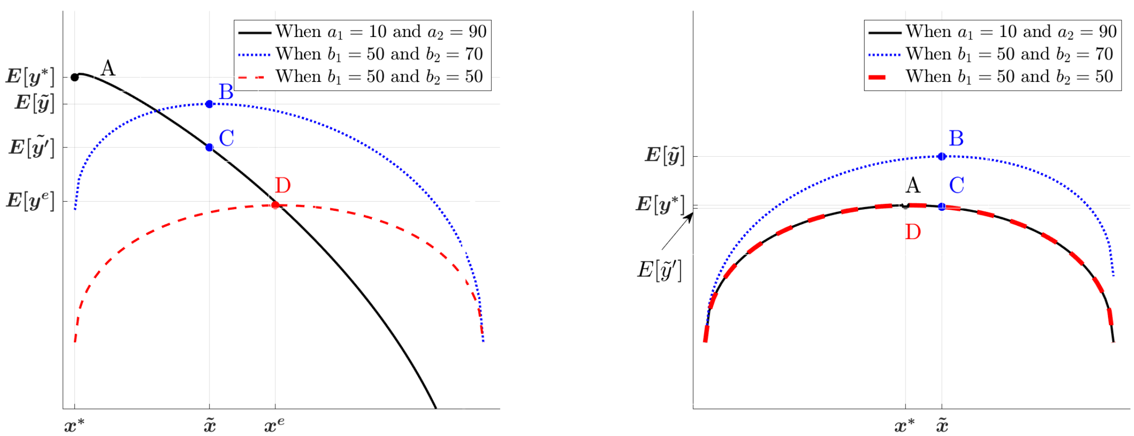
$$\mu^{ITA}(x|a_1, a_2) = a_1\sqrt{x} + a_2\sqrt{100 - x} \tag{1}$$

- **Group Task Assignment (GTA)**. μ depends on the average of a_1 and a_2. The DM must acquire knowledge of this average to make the optimal action.

$$\mu^{GTA}(x|a_1, a_2) = \frac{a_1 + a_2}{2}\sqrt{x} + 50\sqrt{100 - x} \tag{2}$$

2.2 Illustrative example

To elaborate stable and biased beliefs, we examine an illustrative example where $a_1 = 10$ and $a_2 = 90$. We assume that the DM holds biased beliefs about a_1. For instance, in the context of teamwork, the DM believes that the productivity of TM1 is at least as high as 50 and assigns no probability to a_1 being 10 or 30, while TM1's true productivity is 10.



(a) Individual Task Assignments (ITA) (b) Group Task Assignments (GTA)

Figure 1: Illustrative example

Note: The figure presents an example to illustrate stable and biased beliefs. The true parameters are set to be as follows: $a_1 = 10$ and $a_2 = 90$. In both panels, the solid black line represents the actual expected output as a function of action x given the true parameters. The blue dotted line represents the perceived expected output when point beliefs about the true parameters are $b_1 = 50$ and $b_1 = 70$. The DM expects to observe the expected output of $E[\tilde{y}]$ by optimizing her action based on her beliefs (Point B). However, in reality, she observes the expected output of $E[\tilde{y}']$ (Point C). The non-zero *Surprise*, the expectation-reality gap, hints her beliefs are biased; $b_1 = 50$ and $b_1 = 70$ are not stable beliefs. The red dashed line represents the perceived expected output when point beliefs about the true parameters are $b_1 = 50$ and $b_1 = 50$. At point D, the DM maximizes her perceived expected output. Moreover, the perceived expected output is equal to the actual expected output $E[y^e]$. Point D represents the optimal action given the stable (and biased) beliefs and the corresponding expected output. Note that the expected output $E[y^e]$ is inefficient in Individual Task Assignment (panel (a)), but efficient in Group Task Assignment (panel (b)).

Panel (a) of Figure 1 shows the case of Individual Task Assignment (ITA). The black solid line represents the expected output function, which constitutes the objective environment: $\mu^{ITA}(x|a_1 = 10, a_2 = 90)$. Point A in the figure represents the optimal action x^* and the highest expected output achievable $E[y^*] = \mu^{ITA}(x^*|a_1 = 10, a_2 = 90)$.

Suppose the DM holds incorrect beliefs: $b_1 = 50$ and $b_2 = 70$. The blue dotted line in the figure represents the DM's perceived expected output function based on these beliefs: $\mu^{ITA}(x|b_1 = 50, b_2 = 70)$. To maximize the expected output conditional on her wrong beliefs, the DM chooses $\tilde{x}$ and expects to observe output $E[\tilde{y}] = \mu^{ITA}(\tilde{x}|b_1 = 50, b_2 = 70)$ on average (Point B). However, in reality, given the chosen action $\tilde{x}$, the actual expected output is $E[\tilde{y}'] = \mu^{ITA}(\tilde{x}|a_1 = 10, a_2 = 90)$ (Point C). The gap between the realization and expectation, i.e., *surprise*, signals to the DM that her beliefs are inaccurate, specifically that her expectation was too high.

In response, she adjusts her beliefs. Because she believes her productivity is above 50, she revises her beliefs about a_2 rather than a_1. The red dashed line in the figure illustrates the perceived expected output when $b_1 = 50$ and $b_2 = 50$: $\mu^{ITA}(x|b_1 = 50, b_2 = 50)$. Given the new beliefs, the DM chooses the (subjectively) optimal action x^{**} and expects to observe output $E[y^e] = \mu^{ITA}(x^e|b_1 = 50, b_2 = 50)$ on average (Point D). Importantly, at Point D, the perceived output function and the actual output function intersect. This alignment indicates that her expectation coincides with reality. Therefore, Point D represents the optimal action given the stable (and biased) beliefs and the corresponding expected output. Notice that the expected output $E[y^e]$ is inefficient. The DM could have achieved a higher output by $E[y^*] - E[y^e]$.

We do the identical exercise for Group Task Assignment (GTA). The black solid line in Panel (b) of Figure 1 represents the expected output function, which constitutes the objective environment: $\mu^{GTA}(x|a_1 = 10, a_2 = 90)$. Point A in the figure represents the optimal action x^* and the highest expected output achievable $E[y^*] = \mu^{GTA}(x^*|a_1 = 10, a_2 = 90)$.

As before, suppose the DM holds incorrect beliefs: $b_1 = 50$ and $b_2 = 70$. The blue dotted line in the figure represents the DM's perceived expected output function based on these beliefs: $\mu^{GTA}(x|b_1 = 50, b_2 = 70)$. To maximize the expected output conditional on her wrong beliefs, the DM chooses $\tilde{x}$ and expects to observe output $E[\tilde{y}] = \mu^{GTA}(\tilde{x}|b_1 = 50, b_2 = 70)$ on average (Point B). However, in reality, given the chosen action $\tilde{x}$, the actual expected output is $E[\tilde{y}'] = \mu^{GTA}(\tilde{x}|a_1 = 10, a_2 = 90)$ (Point C). The gap between the realization and expectation, i.e., *surprise*, signals to the DM that her beliefs are inaccurate, specifically that her expectation was too high.

In response, she adjusts her beliefs. Because she believes her productivity is above 50, she revises her beliefs about a_2 rather than a_1. The red dashed line in the figure illustrates the perceived expected output when $b_1 = 50$ and $b_2 = 50$: $\mu^{GTA}(x|b_1 = 50, b_2 = 50)$. Given the new beliefs, the DM chooses the (subjectively) optimal action $x^e = x^*$ and expects to observe output $E[y^e] = \mu^{GTA}(x^e|b_1 = 50, b_2 = 50)$ on average (Point D). Importantly, at Point D, the perceived output function and the actual output function intersect. This alignment indicates that her expectation coincides with reality. Therefore, Point D represents the optimal action given the stable (and biased) beliefs and the corresponding expected output. Crucially, the expected output $E[y^{**}]$ is the highest possible output the DM could achieve. In fact, the perceived expected output aligns exactly with the actual expected output, indicating that the DM's understanding of the environment is correct. This is because the DM holds accurate beliefs about the average of a_1 and a_2 even though she has incorrect beliefs about each individually.

2.3 Prediction

As the illustrative example suggests, biased beliefs about a_1 can lead to distinct long-term outcomes, depending on whether it is Individual Task Assignment (ITA) or Group Task Assignment (GTA). This is the core idea of Proposition 5 in Heidhues et al. (2018).

In Individual Task Assignment (ITA), the optimal action requires learning the ratio of a_1 to a_2. If $b_1 \neq a_1$, the ratio of b_1 to a_2 cannot be equal to the ratio of a_1 to a_2 unless $a_1 = a_2$.[3] In short, biased beliefs about a_1 hinder the learning of the a_1/a_2 ratio, resulting in suboptimal actions and inefficient long-term output if $a_1 \neq a_2$.

On the contrary, in Group Task Assignment (GTA), the optimal action requires learning the average of a_1 to a_2. Hence, biased beliefs about a_1 can be compensated by biased beliefs about a_2, not hurting the learning of the average of a_1 and a_2. For example, in the previous example, the stable beliefs $b_1 = b_2 = 50$ overestimate $a_1 = 10$ by 40 and underestimate $a_2 = 90$ by 40. The DM holds the correct beliefs about the average of a_1 and a_2 because the belief biases offset each other. As a consequence, the DM can achieve optimal action and efficient output.

For simplicity, we have illustrated beliefs b_1 and b_2 as point beliefs. Suppose the DM forms belief distributions about a_1 and a_2, and b_1 and b_2 denote the expected value of each belief distribution. Proposition 1 characterizes equilibrium under Individual Task Assignment (ITA) and Group Task Assignment (GTA). Since one can always find belief distributions such that $b_1 \in [10, 90]$ and $b_2 \in [10, 90]$, the first part of the proposition implies that there always exists at least a pair of stable beliefs, except when $a_1 = a_2 = 10$ or when $a_1 = a_2 = 90$. The second part highlights that the optimal action given stable beliefs, called the stable

[3] For instance, when $a_1 = a_2 = 50$, the belief ratio is equal to the true parameter ratio for every pair $b_1 = b_2$.

action) is always optimal; the stable action maximizes the actual expected output in GTA. However, in ITA, the stable action with biased beliefs is optimal only when the ratio of a_1 to a_2 is identical to the belief ratio.

Proposition 1. Let b_1 denote the expected value of the belief distribution about a_1. Let b_2 denote the expected value of the belief distribution about a_2. Suppose the DM optimizes her action conditional on her beliefs and maximizes her perceived expected output.

i. In Individual Task Assignment (ITA), stable beliefs b_1 and b_2 satisfies $(b_1 - \frac{a_1}{2})^2 + (b_2 - \frac{a_2}{2})^2 = \frac{a_1^2+a_2^2}{4}$. The corresponding optimal action x^e maximizes the actual expected output if beliefs about the ratio of a_1 and a_2 are correct, i.e., $\frac{a_1}{a_2} = \frac{b_1}{b_2}$.

ii. In Group Task Assignment (GTA), stable beliefs b_1 and b_2 satisfies $b_1 + b_2 = a_1 + a_2$. The corresponding optimal action x^e maximizes the actual expected output

Proof. See Appendix 5 □

3 Experimental Design

We create a teamwork scenario with two team members, where the productivity of Team Member 1 (TM1) is represented by a_1, and the productivity of Team Member 2 (TM2) is represented by a_2. In the experiments, each subject faces a single-agent decision problem with these two parameters. Throughout the experiment, each subject encounters a fixed set of values for both a_1 and a_2.

We induce belief biases in two ways. First, we rely on naturally occurring belief biases. In Experiment A, each subject is assigned the role of TM1 and is paired with another subject referred to as TM2. Based on extensive research on the better-than-average Zell et al. (2020), we conjecture that subjects with below-average productivity will likely hold incorrect and persistent beliefs about TM1 (themselves).

In Experiment B, we introduce random variations in belief biases. Each subject is paired with two others, one labeled as TM1 and the other as TM2. Subjects receive an informative but noisy signal about TM1's productivity. Regardless of subjects' prior beliefs about TM1, this procedure introduces exogenous belief changes, allowing us to establish a causal link. We anticipate that subjects who receive incorrect signals form more biased beliefs about TM1.

Both Experiment A and Experiment B have two treatments: Individual Task Assignment (ITA) and Group Task Assignment (GTA). The two mutually exclusive task assignment rules differ on whether a subject's optimal decision requires knowledge of each teammate's productivity (ITA) or the average productivity of both team members (GTA). When we need to distinguish the treatments in Experiment B, we refer to them as Egoless Individual Task Assignment (Egoless ITA) and Egoless Group Task Assignment (Egoless GTA). Figure 2 provides an overview of the design.

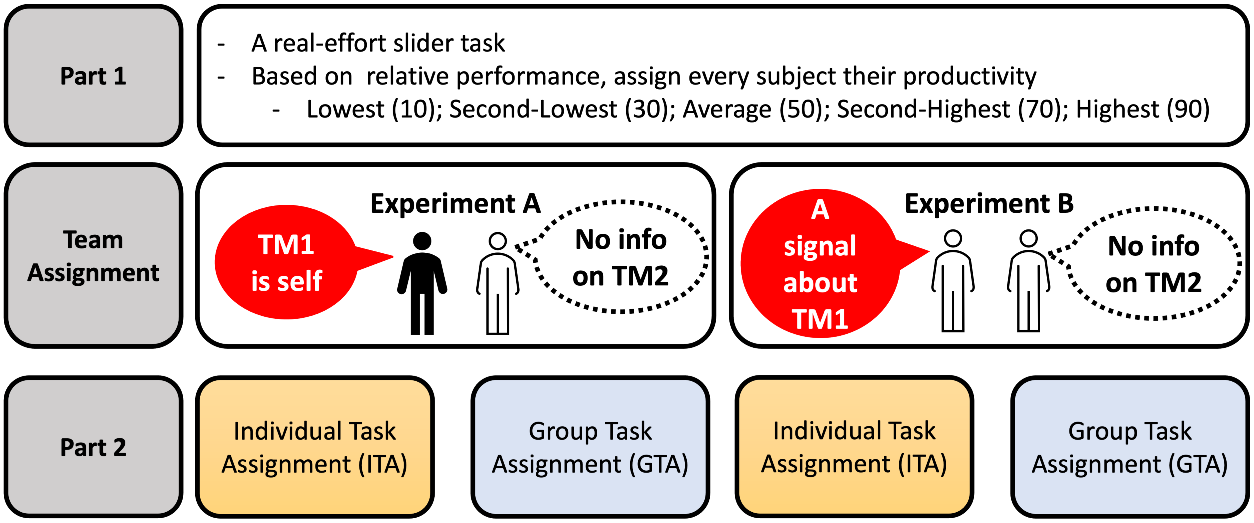


Figure 2: Overview of Experimental Design

3.1 Part 1

In Part 1, subjects complete a modified real-effort slider task (Gill and Prowse, 2012; Gill and Prowse, 2019; Chen and Schildberg-Hörisch, 2019). A single computer screen displays 100 sliders, each scaled from 0 (the very left) to 100 (the very right). Subjects are asked to position as many of the 100 sliders at the center of the respective scale within 5 minutes. Subjects cannot see the numerical position of each slider but can only guess the center position by eye-balling. To move each slider, subjects have to press the left mouse button. The arrows on the keyboard and the mouse wheel are disabled. A slider will count as correctly positioned at the middle position if subjects have set it between 49 and 51, so either exactly at the middle position of 50 or within one position on 50.

Subjects earn 1 Experimental Dollar (ED) for every correctly positioned slider. Furthermore, the instructions clarify: "Your performance on the slider task will also impact your earnings later in the experiment." Subjects receive no information about the performance until the end of the experiment.

At the end of Part 1, every subject is assigned a productivity parameter based on their relative performance. To ensure the independence of each subject's productivity from others, each subject's (absolute) performance in Part 1 is compared with the performance of a fixed group of participants who completed the same slider task before the conduct of this study.[4] If a subject ranks below the 20^{th} percentile (the lowest), her productivity is set to 10. If a subject ranks in the 20^{th}–40^{th} percentile, her productivity is set to 30. If a subject ranks in the 40^{th}–60^{th} percentile, her productivity is set to 50. If a subject ranks in the 60^{th}–80^{th} percentile, her productivity is set to 70. If a subject ranks above the 80^{th} percentile (the highest), her productivity is set to 90.

In the paper, we refer to subjects with a productivity of 10 as the *Lowest* type, those with a productivity of 30 as the *Second-Lowest* type, those with a productivity of 50 as the *Average* type, those with a productivity of 70 as the *Second-Highest* type, and those with a productivity of 90 as the *Highest* type.

3.2 Team assignment

In Experiment A, each subject is randomly paired with another subject. The subject self is designated as TM1, and the paired subject is labeled as TM2.

In Experiment B, every subject is randomly matched with two other participants. Subjects receive a signal about only one of them, referred to as TM1. The signal indicates either "The productivity of TM1 is below 50" or "The productivity of TM1 is above or equal to 50." This signal is true with a 75% chance. No information is provided regarding TM2.

Team assignments remain anonymous and constant in both Experiment A and Experiment B. Every subject encounters one fixed set of parameters throughout the experiment: the productivity of TM1 (a_1) and the productivity of TM2 (a_2).

3.3 Part 2

Part 2 consists of 50 rounds. In each round, subjects complete Task A and Task B. Task A measures subjects' beliefs about TM1 and TM2. Subjects select a point belief (the best guess) from a dropdown menu with options of 10, 30, 50, 70, and 90 for each of a_1 and a_2. Task B asks subjects to state how to allocate 100 hypothetical projects. Subjects can assign only an integer number of projects. Our experiments are a between-subjects design.

In Individual Task Assignment (ITA), each team member's productivity independently determines the expected output, as shown in Equation (1). The contribution of each member to the expected output is proportionate to the number of projects they have been assigned and their productivity. Subjects are instructed to allocate 100 hypothetical projects between

[4] 222 subjects completed the same slider task in Fall 2022.

TM1 and TM2. Subjects observe the realized output, the sum of the expected output, and a random error $\varepsilon \sim N(0, 100)$. Although the chance of encountering a negative random shock exists, the expected output is set high enough to prevent subjects from observing the negative realized output.

In Group Task Assignment (GTA), each team member's productivity jointly determines the expected output. Crucially, the average of their productivity, not each productivity separately, affects the expected output. Subjects are told to allocate 100 hypothetical projects between a group of TM1 and TM2, and the robot player. The productivity of the robot is set to be 50, as shown in Equation (6). Subjects observe the realized output, the sum of the expected output, and a random error $\varepsilon \sim N(0, 100)$. Although the chance of encountering a negative random shock exists, the expected output is set high enough to prevent subjects from observing the negative realized output.[5]

To ensure that subjects form accurate expectations based on their beliefs, we provide a simulator. The simulator displays the expected output conditional on the beliefs reported in Task A. Subjects can experiment with the simulator as many times as they like by entering different numbers in Task A. However, only the final inputs are recorded as the response in Task A. In Group Task Assignment (GTA), subjects are not explicitly told the robot's productivity of 50. Yet, subjects are able to infer it from the simulator. In order to explore the robustness of the results, two additional versions of GTA are conducted in Experiment A with the robot's productivity being 30 and being 70[6]. When we present the results, we pool the three versions of GTA, as the results are robust. Moreover, we offer a history box that enables subjects to track their previous beliefs, actions, and realized outputs.

Both Task A and Task B are incentivized. In Task A, the computer randomly selects one round out of 50. If a subject's beliefs about TM1 and TM2 for the selected round are both correct, they earn 200 ED. If only one of them is correct, they earn 100 ED. If neither guess is correct, they earn 0 ED.

Regarding the payment for Task B, the computer randomly selects one round out of 50, independently of the draw for Task A payment. The computer generates a random number from the interval 0 to 2,500. If the realized output for the chosen round is greater than or equal to the random number, subjects earn 500 ED. Otherwise, they earn 0 ED for Task B. This binary lottery mechanism ensures incentive compatibility regardless of risk preferences

[5] In our data, no subject experiences negative realized output.

[6] Therefore, the expected output function in Group Task Assignment for Experiment A is as follows: c is a constant, either 30, 50, or 70.

$$\mu^{GTA}(x|a_1, a_2) = \frac{a_1 + a_2}{2}\sqrt{x} + c\sqrt{100 - x}$$

Table 1: Treatment conditions

	Individual Task Assignment (ITA)	Group Task Assignment (GTA)
Experiment A	102	206*
Experiment B	99	93

*62 with the robot of 30; 100 with the robot of 50; 56 with robot of 70

(Berg, Daley, Dickhaut, and O'Brien (1986)). Subjects must maximize output to secure the best chance of earning 500 ED.

3.4 Procedure

The experiment was conducted using oTree (Chen, Schonger, and Wickens (2016)) at Texas A&M University. Subjects were recruited using ORSEE (Greiner (2015)).

Table 1 summarizes the treatment conditions. In Experiment A, there are 102 subjects in Individual Task Assignment and 206 subjects in Group Task Assignment. Among the subjects in Group Task Assignment, 62 were assigned to a robot with a productivity of $c =$30, 100 to a robot with a productivity of $c =$50, and 56 to a robot with a productivity of $c =$70. In Experiment B, 99 subjects participated in Egoless Individual Task Assignment, and 93 subjects participated in Egoless Group Task Assignment with a robot of productivity $c =$50. The experiment lasted, on average, one hour. Payments averaged approximately 15.23.

Once subjects are seated in a computer station, an experimenter read aloud the instructions for Part 1. While completing Part 1, subjects know the existence of Part 2, but do not know what the task will be. After Part 1 ends, the experimenter distributes the instructions for the rest of the experiment and read them aloud. Subjects enter Part 2 at the same time, but they make decisions at their own pace. To prevent rushed decisions as much as possible, every subject must remain seated until a session ends.

3.5 Discussion

Proposition 1 suggests that a pair of stable beliefs exists for all pairs (a_1, a_2) except for (10, 10) and (90, 90) under the assumptions of probabilistic beliefs and continuous action between 0 and 100. However, it is infeasible to elicit belief distributions for 50 rounds of experiments due to time constraints. Moreover, there is evidence that cognitive complexity and burden negatively impact data quality. Therefore, we restrict subjects to reporting point beliefs (their best guesses) as one of five possible values: 10, 30, 50, 70, or 90. We also allow subjects to choose only whole numbers ranging from 0 to 100. Appendix 5 presents all possible pairs of stable, and biased, beliefs in Individual Task Assignment. There exists at least a pair of stable beliefs in Group Task Assignment for every pair (a_1, a_2) but for (10, 10) and (90, 90).

3.6 Hypotheses

Individual Task Assignment (ITA) and Group Task Assignment (GTA) require distinct knowledge for optimal decision-making. We hypothesize that learning will be affected by these incentives. We formalize this intuition in a series of hypotheses.

Hypothesis 1.1. (Learning the productivity ratio) The ratio between beliefs about TM1's productivity and beliefs about TM2's productivity is closer to the true productivity ratio of TM1 and TM2 in ITA than GTA.

Hypothesis 1.2. (Learning the average productivity) The average of beliefs about the productivity of TM1 and TM2 is closer to the true average productivity of TM1 and TM2 in GTA than in ITA.

Experiment A relies on a well-documented tendency that people (mistakenly) believe they are better than the average, especially on easy tasks (e.g., Moore and Healy, 2008; Svenson, 1981; Zell et al., 2020). This better-than-average effect suggests that all subjects would believe their performance of the trivial real-effort task in Part 1 is at least as high as the average. Subjects, especially whose performance is truly below the average, would tend to overestimate their relative performance and assign only a slim chance to being below average. Consequently, we anticipate that *Lowest* and *Second-Lowest* types are more likely to exhibit biased beliefs about TM1 (self).

In Experiment B, subjects make allocation decisions based on their beliefs about two other team members, TM1 and TM2. Beliefs about TM1 are manipulated by a noisy yet informative signal. Subjects who receive an incorrect signal about TM1 are more likely to develop and maintain inaccurate beliefs. If a subject receives a signal indicating that TM1's productivity is above or equal to 50, but in reality, TM1's productivity is less than 50, then the subject may become fixated on the signal value and have a harder time learning TM1's true productivity. By integrating these conjectures with Proposition 5 from Heidhues et al. (2018), we arrive at Hypothesis 2.

Hypothesis 2.1. (Treatment effects in Experiment A) In Experiment A, *Lowest* and *Second-Lowest* experience greater inefficiency in ITA compared to the other types, although this difference is not observed in GTA.

Hypothesis 2.2. (Treatment effects in Experiment B) In Experiment B, receiving incorrect signals about TM1 causes greater inefficiency in ITA. However, the inefficiency is disappears in GTA.

Lastly, we examine whether beliefs converge to stable beliefs. According to definition 1, the DM holding stable beliefs experiences no gap between the perceived expected output and the actual expected output; *Surprise* is zero. Since *Surprise* can be positive or negative depending on whether subjects' expectation are too high or too low, we use the absolute value of *Surprise* in the analysis: $\textit{Abs. Surprise} = |\mu(x|a_1, a_2) - \mu(x|b_1, b_2)|$. As beliefs converge to stable beliefs, *Abs. Surprise* should approach zero.

An empirical question related is how subjects interpret *Abs. Surprise*. Consider two subjects: one subject perceives an expected output of 100 while her actual expected output is 90, and another subject perceives an expected output of 1000 while her actual expected output is 990. Although the *Abs. Surprise* for both subjects is 10, they may interpret it differently. The former observes 10% less expected output than what she anticipates, while the latter observes only a 1% difference between her perceived and actual expected output. Therefore, we capture the degree by which subjects perceive expectation-reality gaps. *Abs. Surprise rel. Reality* is the proportion of *Abs. Surprise* relative to their perceived expected output. We expect that subjects with larger *Abs. Surprise rel. Reality* converge to stable beliefs faster than those with smaller values as they perceive a more salient discrepancy between expectation and reality (e.g., Bordalo, Gennaioli, and Shleifer, 2012)

$$\textit{Abs. Surprise rel. Reality} = \frac{|\mu(x|a_1, a_2) - \mu(x|b_1, b_2)|}{\mu(x|a_1, a_2)}$$

Hypothesis 3.1. (Actual expectation-reality gap) *Abs. Surprise* decrease over time and approaches zero.

Hypothesis 3.2. (Perceived expectation-reality gap) Subjects with higher initial values of *Abs. Surprise rel. Expectation* experience a faster decline in this metric.

4 Results

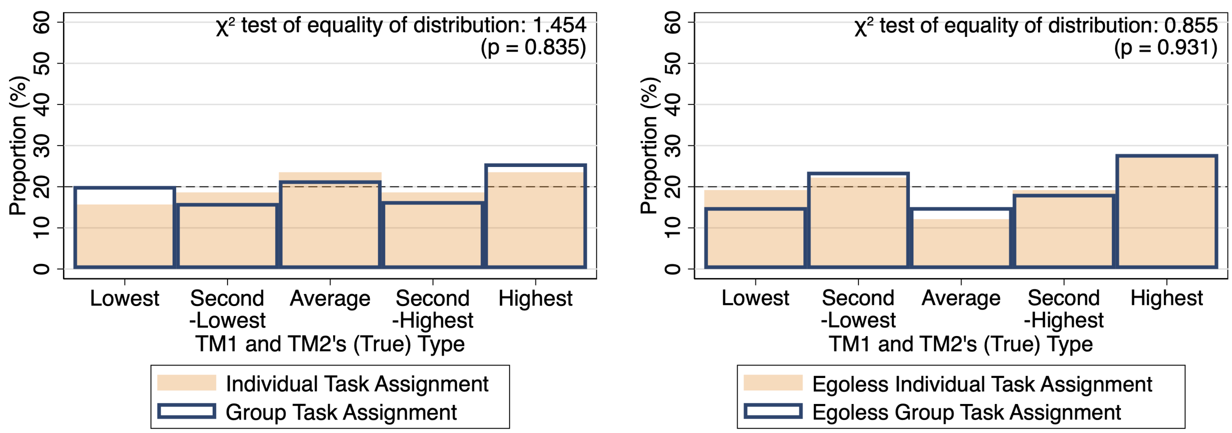
We present our main results as follows: In Section 4.1, we confirm that subjects adjust their beliefs with different objectives across treatments, which is consistent with Hypothesis 1. In Section 4.2, we document different outcomes between Individual Task Assignment (ITA) and Group Task Assignment (GTA). We find biased beliefs lead to greater inefficiency in ITA than in GTA, in line with Hypothesis 2. Section 4.3 provides little evidence to support Hypothesis 3, as subjects' beliefs consistently mismatch reality.

4.1 Learning patterns

Section 4.1.1 assesses balanced treatment assignments. We show that there is no systematic difference in beliefs between Individual Task Assignment (ITA) and Group Task Assignment (GTA) both in Experiment A and Experiment B. Section 4.1.2 presents evidence supporting Hypothesis 1. This section also highlights that *Lowest* type subjects exhibit the most biased beliefs in Experiment A, while subjects receiving incorrect signals about TM have more biased beliefs than those receiving correct signals in Experiment B. In Section 4.1.3, we show that subjects optimize their choices based on their beliefs.

4.1.1 Balanced treatment assignment

We test for differences in true productivity distributions between treatments.



(a) Experiment A (b) Experiment B

Figure 3: Histograms of (true) productivity

Note: Panel (a) presents the histogram of the subjects' (true) productivity in Experiment A. The true distributions are not statistically different from the uniform distribution in both Individual Task Assignments (Pearson $\chi^2 = 2.412$, $p = 0.661$) and Group Task Assignments (Pearson $\chi^2 = 6.817$, $p = 0.146$). Furthermore, there are no statistically significant differences between treatments (the statistics are reported in the figures). Panel (b) presents the histogram of the subjects' (true) productivity in Experiment B. The true distributions are not statistically different from the uniform distribution in both in Egoless Individual Task Assignment (Egoless ITA) (Pearson χ^2 for Egoless ITA: 6.0, $p = 0.199$) and Egoless Group Task Assignment (Egoless GTA) (Pearson $\chi^2 = 5.978$, $p = 0.201$). Furthermore, there are no statistically significant differences between treatments (the statistics are reported in the figures).

Figure 3a shows histograms of true productivity for Individual Task Assignment (ITA) and Group Task Assignment (GTA) in Experiment A. The distributions in both conditions are not statistically different from a uniform distribution (Pearson $\chi^2 = 2.412$, $p = 0.661$ for ITA; Pearson $\chi^2 = 6.817$, $p = 0.146$ for GTA). The two distributions are not significantly different ($\chi^2 = 1.454$, $p = 0.835$).

Figure 3b compares the true productivity distributions between Egoless Individual Task Assignment (Egoless ITA) and Egoless Group Task Assignment (Egoless GTA) in Experiment B. The true distributions are not statistically different from the uniform distribution (Pearson χ^2 for Egoless ITA: 6.0, $p = 0.199$; Pearson χ^2 for Egoless GTA: 5.978, $p = 0.201$). Furthermore, there are no statistically significant differences between treatments (χ^2 =0.855, p =0.931).

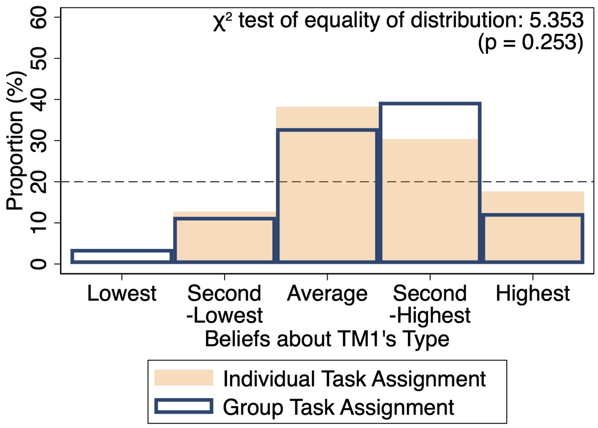
Next, we analyze the distributions of beliefs about TM1 and TM2. Figure 4 presents belief distributions in Experiment A. Figure 5 presents belief distributions in Experiment B. Recall that in Experiment A, TM1 represents the subject themselves whereas TM1 is a randomly-selected participant in Experiment B.

Belief distributions in Experiment A

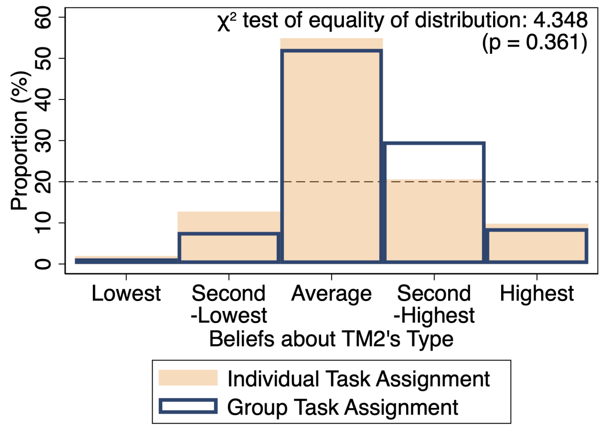
Figure 4a are histograms of beliefs about TM1 (self) in the initial round of Experiment A. In Individual Task Assignment (ITA), less than 1% of subjects report TM1's productivity as the *Lowest*, and 12.8% report it as the *Second-Lowest*. These proportions are lower than the actual distribution: *Lowest* and *Second-Lowest* types account for 15.7% and 18.6% of our sample, respectively. In Group Task Assignment (GTA), 15.1% (3.7% and 11.5% of subjects) report that TM1 is below average, while in reality, 36.2% (20.2% and 16.1%) are below average. Unlike the distribution of true productivity, the belief distributions significantly differ from the uniform distribution (Pearson χ^2 in ITA: 43.882, $p = 0.000$; Pearson χ^2 in GTA: 103.055, $p = 0.000$), but they are not significantly different between treatments (χ^2 test of equality of distribution: 5.353, $p = 0.253$).

Figure 4b presents the histograms of reported beliefs about TM2. Subjects also assign a slim chance to the teammate being the *Lowest* or the *Second-Lowest* in round 1. Most subjects initially report TM2's productivity as the *Average*. In Individual Task Assignment (ITA), 85.3% report that TM2's productivity is as high as the average, while in reality, it is 65.7%. In Group Task Assignment (GTA), 90.8% report that TM2's productivity is as high as the average, while in reality, it is 63.8%. The belief distributions are statistically different from the uniform distribution (Pearson χ^2 in ITA: 86.725, $p = 0.000$; Pearson χ^2 in GTA: 192.092, $p = 0.000$). They are not different across treatments (χ^2 test of equality of distribution: 4.348, $p = 0.361$).

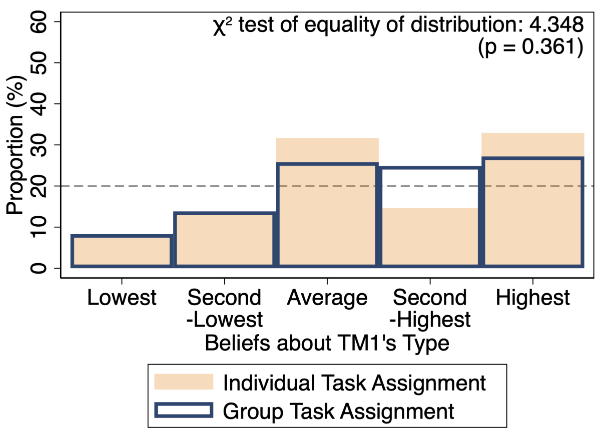
Figure 4c and 4d present belief distributions in the last round. The clustering at the *Average* and *Second-Highest* productivity is less pronounced compared to the histograms in Round 1 (Figure 4a and 4b). The proportion of subjects reporting TM1 (self) as being below average increases to 20.7% in Individual Task Assignment (ITA) and 22.1% in Group



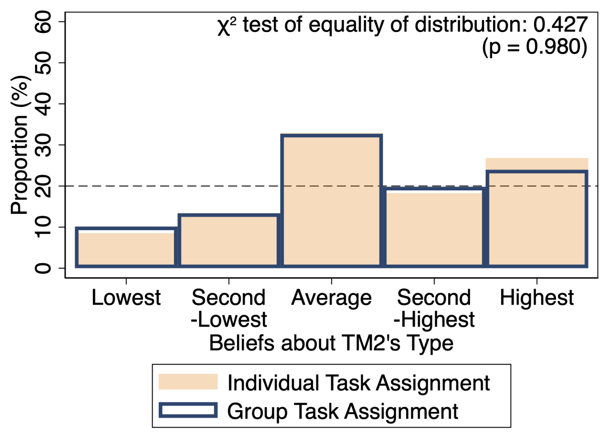
(a) Beliefs about TM1 (self) in Round 1



(b) Beliefs about TM2 in Round 1



(c) Beliefs about TM1 (self) in Round 50



(d) Beliefs about TM2 in Round 50

Figure 4: Histograms of beliefs in Experiment A

Note: Panels (a) and (b) show histograms of beliefs about the productivity of TM1 (self) and TM2 in the first round. The belief distributions about TM1 (self) are statistically different from the uniform distribution in Individual Task Assignments (ITA) (Pearson χ^2: 43.882, $p = 0.000$) and Group Task Assignments (GTA) (Pearson $\chi^2 = 103.055$, $p = 0.000$). The belief distributions about TM2 are also statistically different from the uniform distribution in ITA (Pearson χ^2: 86.725, $p = 0.000$) and GTA (Pearson $\chi^2 = 192.092$, $p = 0.000$). Panels (c) and (d) display the same histograms in the last round. All belief distributions are statistically different from the uniform distribution. For beliefs about TM1 (self), Pearson χ^2 in ITA: 22.024, $p = 0.000$; Pearson χ^2 in GTA: 30.857, $p = 0.000$. For beliefs about TM2, Pearson χ^2 in ITA: 16.049, $p = 0.003$; Pearson χ^2 in GTA: 34.59, $p = 0.000$. The distributions undergo significant changes from Round 1 to Round 50. Concerning beliefs about TM1 (self), the χ^2 test of equality of distribution in ITA is 14.531, $p = 0.006$, and in GTA it is 25.52, $p = 0.000$. For beliefs about TM2, the χ^2 test of equality of distribution in ITA is 16.599, $p = 0.002$; and in GTA it is 47.382, $p = 0.000$). Across all distributions, we do not find differences between treatments (The statistics are reported in the figures).

Task Assignment (GTA). The proportion of subjects reporting TM2 as below average also increases to 22.0% in ITA and 23.5% in GTA. However, these proportions are still substantially smaller than the true proportions.

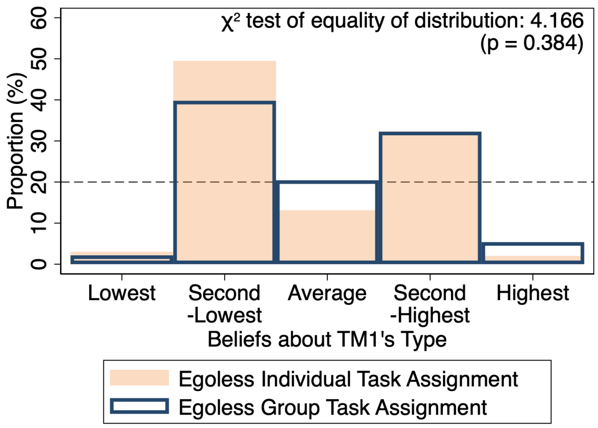
The changes in belief distributions from Round 1 to Round 50 are statistically significant. For TM1 (self), the χ^2 test of equality of distribution in Individual Task Assignment (ITA) is 14.531 ($p = 0.006$), while in Group Task Assignment (GTA), it results in 25.52 ($p = 0.000$). Regarding TM2, the χ^2 test of equality of distribution in ITA is 16.599 ($p = 0.002$), and in GTA, it results in 47.382 ($p = 0.000$). However, we still find no difference in belief distributions between treatments for TM1 (self) (χ^2 test of equality of distribution: 4.348, $p = 0.361$) and for TM2 (χ^2 test of equality of distribution: 0.427, $p = 0.980$).

Belief distributions in Experiment B

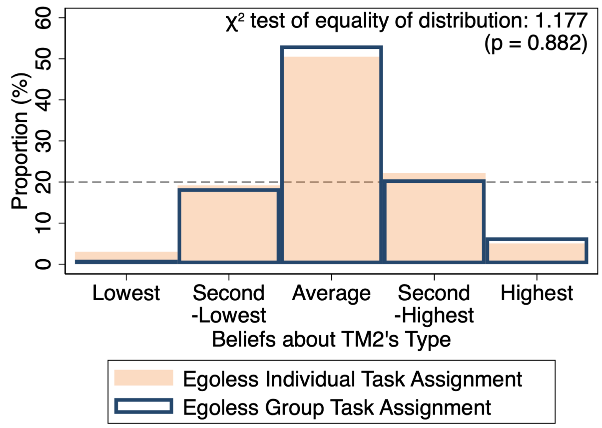
Figure 5a and 5b show the belief distributions of TM1' and TM2' productivities in the first round of Experiment B. When comparing these distributions to those in Experiment A, the overall shape suggests that when TM1 is a random person instead of the subject themselves, a greater proportion of subjects report the *Second-Lowest* productivity as a possibility. However, very few subjects report TM1 as the *Lowest*. The distributions are different from the uniform distribution. Beliefs about TM1 in round 1: Pearson χ^2 in Egoless ITA: 83.172 ($p = 0.000$); Pearson χ^2 in Egoless GTA: 49.957 ($p = 0.000$). Beliefs about TM2 in round 1: Pearson χ^2 in Egoless ITA: 71.657 ($p = 0.000$); Pearson χ^2 in Egoless GTA: 75.826 ($p = 0.000$).

Figure 5c and 5d show the belief distributions in the last round. In Egoless Individual Task Assignment (Egoless ITA), the distributions are not statistically different from the uniform distribution. For beliefs about TM1, Pearson χ^2 in Egoless ITA is 7.278 ($p = 0.122$). For beliefs about TM2, Pearson χ^2 in Egoless ITA is 3.876 ($p = 0.423$). In the absence of ego, subjects perform better at learning the productivity of team members. Conversely, in Egoless Group Task Assignment (Egoless GTA), the belief distributions are statistically different from the uniform distribution. For beliefs on TM1, Pearson χ^2 in Egoless GTA is 15.333 ($p = 0.004$). For beliefs about TM2 ($p = 0.423$), Pearson χ^2 in Egoless GTA is 17.783 ($p = 0.001$).

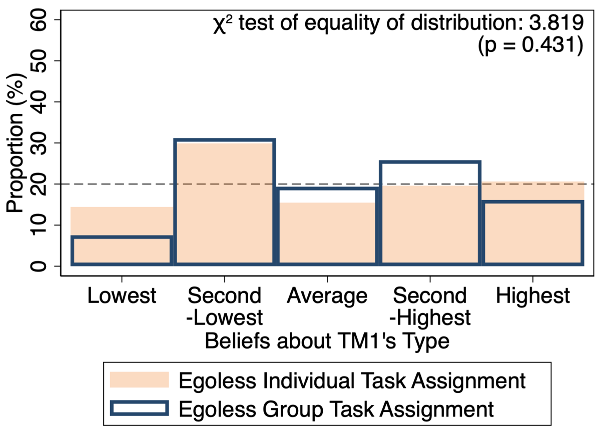
In both treatments, the belief distributions significantly change from round 1 to round 50, indicating learning. Regarding TM1, the χ^2 test of equality of distribution between round 1 and round 50 is 30.412 ($p = 0.000$) in Egoless ITA, and 9.441 ($p = 0.051$) in Egoless GTA. Regarding TM2, the χ^2 test of equality of distribution is 29.783 ($p = 0.000$) in Egoless ITA, and 18.706 ($p = 0.001$) in Egoless GTA.



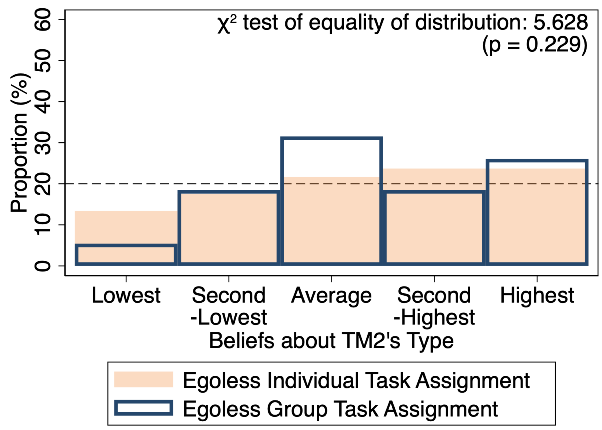
(a) Beliefs about TM1 in Round 1



(b) Beliefs about TM2 in Round 1



(c) Beliefs about TM1 in Round 50



(d) Beliefs about TM2 in Round 50

Figure 5: Histograms of beliefs in Experiment B

Note: Panels (a) and (b) present the histograms of beliefs about the productivity of TM1 and TM2 in the first round. The belief distributions about TM1 are statistically different from the uniform distribution in Egoless Individual Task Assignment (Egoless ITA) (Pearson χ^2: 83.172, $p = 0.000$) and Egoless Group Task Assignment (Egoless GTA) (Pearson χ^2: 49.957, $p = 0.000$). The belief distributions about TM2 are also statistically different from the uniform distribution in Egoless ITA (Pearson χ^2: 71.657, $p = 0.000$) and Egoless GTA (Pearson χ^2: 75.826, $p = 0.000$). Panels (c) and (d) present the same histograms in the last round. Belief distributions in Egoless ITA are not statistically different from the uniform distribution. For beliefs about TM1, Pearson χ^2 in Egoless ITA: 7.278 ($p = 0.122$). For beliefs about TM2, Pearson χ^2 in Egoless ITA: 3.876 ($p = 0.423$). However, belief distributions in Egoless GTA are statistically different from the uniform distribution. For beliefs about TM1, Pearson χ^2 in Egoless GTA: 15.333 ($p = 0.004$). For beliefs about TM2, Pearson χ^2 in Egoless GTA: 17.783 ($p = 0.001$). The belief distributions undergo significant changes from round 1 to round 50. Concerning beliefs about TM1, the χ^2 test of equality of distribution in Egoless ITA is 30.412 ($p = 0.000$) and in Egoless GTA it is 9.441 ($p = 0.051$). The beliefs about TM2, the χ^2 test of equality of distribution in Egoless ITA is 29.783 ($p = 0.000$) and it is 18.706, ($p = 0.001$) in Egoless GTA. Across all distributions, we do not find differences between treatments (The statistics are reported in the figures).

Finally, no significant differences are found between Egoless Individual Task Assignment (Egoless ITA) and Egoless Group Task Assignment (Egoless GTA) across all distributions. In round 1, the χ^2 test for equality of distribution between Egoless ITA and Egoless GTA: 4.166 (p = 0.384) for TM1 and 1.177 (p = 0.882) for TM2. In round 50, the χ^2 test for equality of distribution: 3.819 (p = 0.431) for TM1 and 5.628 (p = 0.229) for TM2.

4.1.2 Learning the productivity ratio and the average productivity

Our treatments impose different learning objectives on subjects. In Individual Task Assignment (ITA), subjects need to learn the productivity of TM1 relative to the productivity of TM2. In Group Task Assignment (GTA), subjects need to learn the average productivity of TM1 and TM2. Therefore, we investigate whether subjects display different learning patterns by treatments in terms of learning the productivity ratio of TM1 to TM2 and learning the average productivity of TM1 and TM2.

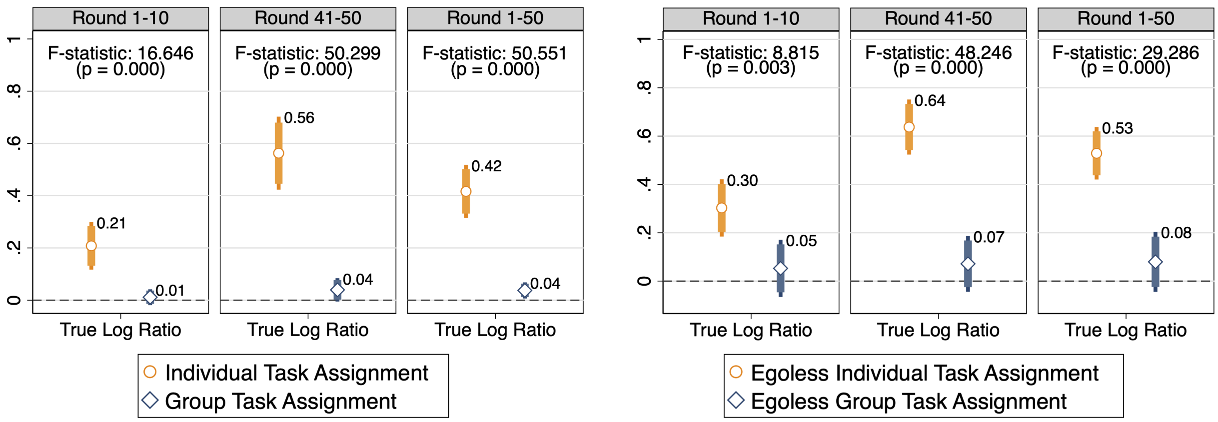
We regress subjects' beliefs (y_{it}) on the true value (y_i^*) as shown in Equation (3), where i represents an individual and $t \in \{1, 2, \cdots, 50\}$ represents a round. If subjects have the correct belief, the coefficient α_1 should be 1. We control for rounds, and standard errors are clustered at the individual level.

$$y_{it} = \alpha_0 + \alpha_1 y_i^* + \epsilon_{it} \tag{3}$$

Learning the productivity ratio

Figure 6 plots the estimated α_1 of Equation (3) when the dependent variable (y_{it}) is *Log Ratio of Beliefs*, defined by the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2. The corresponding true ratio is (y_i^*), referred to as the *True Log Ratio*, is the log ratio of the true productivity of TM1 and the true productivity of TM2. Table A.1 and A.2 report the full regression estimates.

Figure 6a presents the results of Experiment A. The left and middle panels show the estimates during the first and last ten rounds, respectively. The right panel pools all 50 rounds. All three panels show that the coefficients of *True Log Ratio* in Individual Task Assignment (ITA) are significantly greater than those in Group Task Assignment (GTA). In ITA, the coefficient starts at around 0.21 in the first ten rounds and increases to almost 0.56 in the last ten rounds. In contrast, in GTA, the coefficients remain around zero throughout. When pooling all rounds, the coefficient in GTA is less than one-tenth of that in ITA. The coefficients are significantly different between ITA and GTA (F-statistic: 50.551, p =0.000).



(a) Experiment A

(b) Experiment B

Figure 6: Learning the productivity ratio

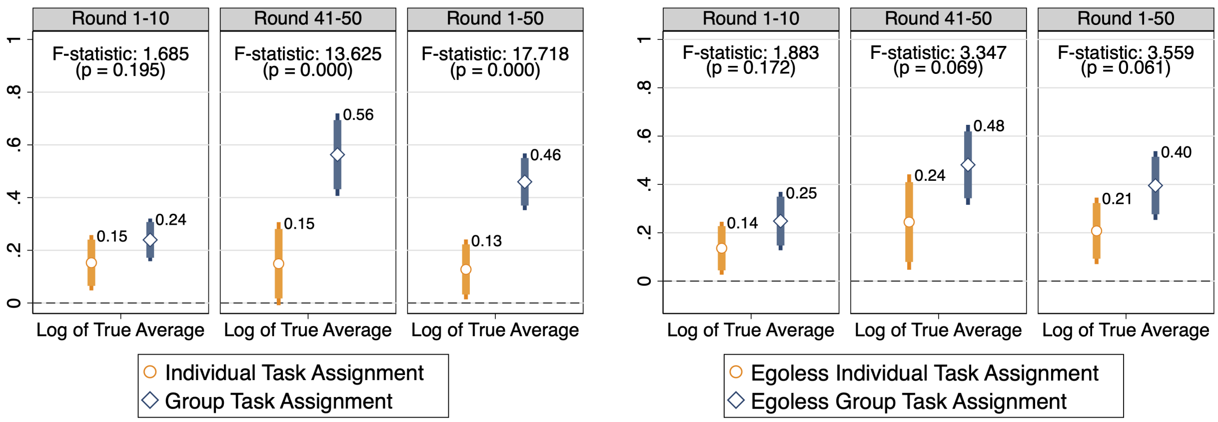
Note: The figures present the estimates of Equation (3). If subjects have the correct belief about the productivity ratio of TM1 and TM2, the coefficient should be 1. Error bars represent 95% and 90% confidence intervals. In all regressions, we control for rounds, and standard errors are clustered at the individual level. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between Individual Task Assignment and Group Task Assignment. Table A.1 and A.2 report the full regression estimates.

Figure 6b presents the results of Experiment B. The figure shows qualitatively identical learning patterns to those observed in Experiment A. Subjects in Egoless Individual Task Assignment (Egoless ITA) have beliefs closer to the productivity ratio between TM1 and TM2 more than subjects in Egoless Individual Group Assignment (Egoless GTA). The coefficient in Egoless ITA starts at around 0.3 in the first ten rounds and increases to almost 0.64 in the last ten rounds. In contrast, the coefficients in Egoless GTA are less than 0.1 during the 50 rounds. When pooling all rounds, the coefficient in Egoless GTA is approximately one-tenth of that in Egoless ITA. The coefficients are significantly different between Egoless ITA and Egoless GTA (F-statistic: 29.286, p =0.000).

Learning the average productivity

Figure 7 plots the estimated α_1 of Equation (3) when the dependent variable (y_{it}) is *Log of Average Beliefs*, defined by the log of beliefs about the average productivity of TM1 and TM2. The corresponding truth (y_i^*), referred to as *Log of True Average*, is the log of the true average productivity of TM1 and TM2. Table A.3 and A.4 report the full regression estimates.

Figure 7a presents the results of Experiment A. The left and middle panels show the estimates during the first and last ten rounds, respectively. The right panel pools all 50



(a) Experiment A

(b) Experiment B

Figure 7: Learning the average productivity

Note: The figures present the estimates of Equation (3). If subjects have the correct belief about the average productivity of TM1 and TM2, the coefficient should be 1. Error bars represent 95% and 90% confidence intervals. In all regressions, we control for rounds, and standard errors are clustered at the individual level. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between Individual Task Assignment and Group Task Assignment. Table A.3 and A.4 report the full regression estimates.

rounds. All three panels show that the coefficients of *Log of True Average* in Individual Task Assignment (ITA) are significantly smaller than those in Group Task Assignment (GTA). In ITA, the coefficient consistently remains below 0.2. On contrast, the coefficient in GTA is 0.24 in the first ten rounds, not significantly different from that in ITA (F-statistic: 1.685, p =0.195). However, it increases to 0.56 and the difference becomes significant in the last ten rounds (F-statistic: 13.625, p =0.000). Pooling all rounds, the coefficient in ITA is almost a quarter of that in GTA. The coefficients are significantly different between ITA and GTA (F-statistic: 17.718, p =0.000).

Figure 7b presents the results of Experiment B. We observe the same learning patterns as in Experiment A. Beliefs deviate further from the average productivity of TM1 and TM2 in Egoless Individual Task Assignment (Egoless ITA) than in Egoless Group Task Assignment (Egoless GTA). In the first ten rounds, subjects in both treatments have equally biased beliefs about the average (F-statistic: 1.883, p =0.172). However, the gap emerges over time. In the last ten rounds, belief biases among subjects in Egoless ITA is twice as large as those among subjects in Egoless GTA (F-statistic: 3.347, p =0.169). Pooling all rounds, the coefficient in Egoless GTA is almost a double of that in Egoless ITA. The coefficients are significantly different between Egoless ITA and Egoless GTA (F-statistic: 3.559, p =0.061).

Who has more biased beliefs?

Now, we investigate who has more or less biased beliefs. In Experiment A, every subject herself is designated as TM1. Based on the better-than-average effect (Zell et al., 2020), we expect *Lowest* and *Second-Lowest* type subjects would hold more biased beliefs. In Experiment B, on the other hand, subjects receive an exogenous signal. Receiving incorrect signals about TM1, regardless of their prior beliefs concerning TM1, is expected to trigger more biased beliefs.

We regress the deviation of beliefs (y_{it}) from the true value (y_i^*) on the indicators of subject types in Experiment A, and the indicator of receiving incorrect signals about TM1 in Experiment B. There are two dependent variables of interest. One is the absolute distance between *Log Ratio of Beliefs*, defined by the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2, and the true productivity ratio. The other is the absolute distance between *Log of Belief Average*, defined by the log of beliefs about the average productivity of TM1 and TM2, and the true average productivity. We control for rounds, and standard errors are clustered at the individual level.

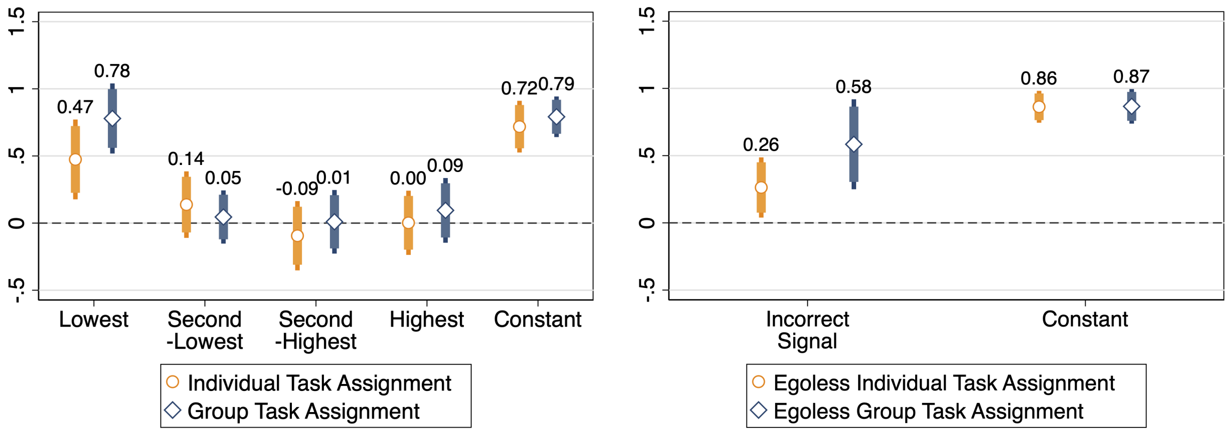
$$|y_{it} - y_i^*| = \beta_0 + \beta_1 Indicator_i + \epsilon_{it} \tag{4}$$

Figure 8 plots the estimated β_1 of Equation (4) regarding belief biases in the productivity ratio between TM1 and TM2 across subject types. Table A.5 and A.6 report the full regression estimates.

Figure 8a presents coefficient plots for Experiment A, pooling all rounds. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional belief biases with respect to the productivity ratio of TM1 and TM2 that *Lowest* types make compared to *Average* types.

As shown in the figure, *Lowest* types in Experiment A exhibit significantly greater belief biases than *Average* types. Beliefs about the productivity ratio of TM1 and TM2 among *Lowest* types deviate further from the true ratio compared to *Average* types by 0.47 in Individual Task Assignment (ITA) and by 0.78 in Group Task Assignment (GTA). However, *Second-Lowest*, *Second-Highest* and *Highest* types do not exhibit more biased beliefs than *Average* types in both treatments.

We also find that, compared to *Average* types, every subject type has no more biased beliefs in Individual Task Assignment (ITA) and Group Task Assignment (GTA). For example, the difference in belief biases between *Lowest* and *Average* types in ITA, 0.47, is not



(a) Experiment A

(b) Experiment B

Figure 8: Biases in beliefs about the productivity ratio

Note: The figures present the estimates of Equation (4). The dependent variable is the absolute distance between *Log Ratio of Beliefs*, defined by the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2, and the true productivity ratio of TM1 and TM2. In panel (a), the coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional belief biases with respect to the productivity ratio of TM1 and TM2 that *Lowest* types have compared to *Average* types. In panel (b), the coefficients indicate the causal effect of receiving incorrect signals about TM1. The coefficient of *Incorrect Signal* captures the additional belief biases resulting from receiving incorrect signals. Error bars represent 95% and 90% confidence intervals. In all regressions, we control for rounds, and standard errors are clustered at the individual level. Table A.5 and A.6 report the full regression estimates.

statistically different from the difference in belief biases between *Lowest* and *Average* types in GTA, 0.778 (F-statistic: 2.338, p =0.127).[7] This confirms the comparability of belief biases between ITA and GTA.

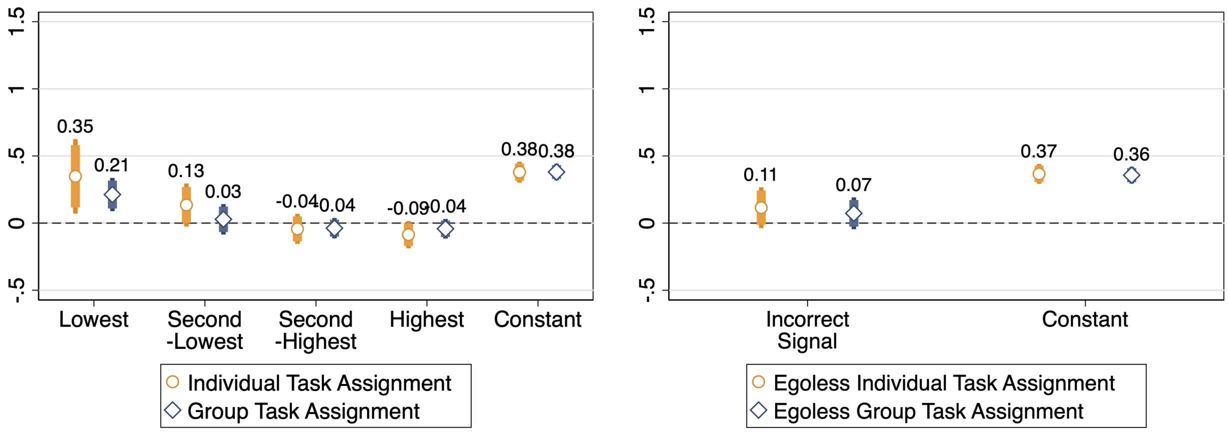
Figure 8b presents coefficient plots for Experiment B, pooling all rounds. The coefficients indicate the causal effect of receiving incorrect signals about TM1. The coefficient of *Incorrect Signal* captures the additional belief biases resulting from receiving incorrect signals.

The figure shows that subjects on average have biases in beliefs about the productivity ratio by 0.86 in Egoless Individual Task Assignment (Egoless ITA) and 0.87 in Egoless Group Task Assignment (Egoless GTA). Receiving incorrect signals about TM1 further increases the biases by 0.26 in Egoless ITA and by 0.58 in Egoless GTA. The significant effect of receiving incorrect signals remains after controlling for potential impacts of TM1's type (See Column (7)-(12) of Table A.6).[8] We also find that the causal effect of receiving incorrect

[7] F-statistic: 0.334 (p =0.564) for the difference in difference between *Second-Lowest* and *Average* types; F-statistic: 0.343 (p =0.559) for the difference in difference between *Second-Highest* and *Average* types; F-statistic: 0.292 (p =0.590) for the difference in difference between *Highest* and *Average* types.

[8] For instance, subjects may hesitate to believe that TM1's productivity is the lowest, just as they might hesitate to believe their own productivity is the lowest, leading to greater belief biases.

signals about TM1 in Egoless ITA, 0.26, is not statistically different from the causal effect in Egoless GTA, 0.58 (F-statistic: 2.518, p =0.114). This confirms the comparability of belief biases between Egoless ITA and Egoless GTA.



(a) Experiment A

(b) Experiment B

Figure 9: Belief biases in the (true) average productivity

Note: The figures present the estimates of Equation (4). The dependent variable is the absolute distance between *Log of Average Beliefs*, defined by the log of beliefs about the average productivity of TM1 and TM2, and the true average productivity. In panel (a), the coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional belief biases with respect to the average productivity of TM1 and TM2 that *Lowest* types have compared to *Average* types. In panel (b), the coefficients indicate the causal effect of receiving incorrect signals about TM1. The coefficient of *Incorrect Signal* captures the additional belief biases resulting from receiving incorrect signals. Error bars represent 95% and 90% confidence intervals. In all regressions, we control for rounds, and standard errors are clustered at the individual level. Table A.7 and A.8 report the full regression estimates.

Figure 9 plots the estimated β_1 of Equation (4) regarding belief biases in the average productivity of TM1 and TM2. Rounds are pooled together in both figures. Table A.7 and A.8 report the full regression estimates. The figure highlights that subjects have less biased beliefs about the average productivity of TM1 and TM2 compared to their belief biases in the productivity ratio between TM1 and TM2.

Figure 9a presents coefficient plots for Experiment A. It shows that *Average* types' belief biases regarding the average productivity of TM1 and TM2 in Individual Task Assignment (ITA) are approximately 0.38. Belief biases of *Lowest* and *Second-Lowest* are greater than *Average* types' by 0.35 and 0.13, respectively. No differences are observed between *Second-Highest* and *Highest* types compared to *Average* types. In Group Task Assignment (GTA), *Average* types' belief biases regarding average productivity are also 0.38. *Lowest* types hold more biased beliefs than *Average* types by 0.21.

These coefficients are smaller when compared to belief biases regarding the productivity ratio between TM1 and TM2 (as shown in Figure 8a). For instance, *Average* types' belief biases regarding the productivity ratio are 0.72 in ITA, which is nearly double their belief biases regarding average productivity. In GTA, *Lowest* types exhibit belief biases in terms of the productivity ratio that are three times greater than those of *Average* types although *Average* types' belief biases regarding the productivity ratio are already twice as large as their belief biases regarding average productivity.

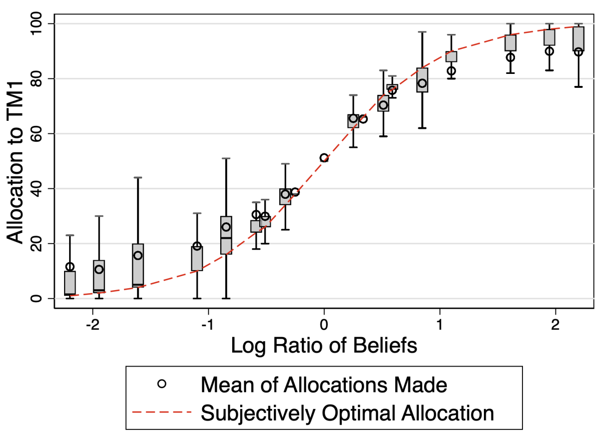
We also find that, compared to *Average* types, every subject type has no more biased beliefs in Individual Task Assignment (ITA) and Group Task Assignment (GTA). For example, the difference in belief biases between *Lowest* and *Average* types in ITA, 0.35, is not statistically different from the difference in belief biases between *Lowest* and *Average* types in GTA, 0.21 (F-statistic: 0.787, p =0.376).[9] This confirms the comparability of belief biases between ITA and GTA.

Figure 9b presents coefficient plots for Experiment B. It strengthens the findings in Experiment A, providing causal evidence. In Experiment B, subjects who receive correct signals about TM1 exhibit biased beliefs about the average productivity of TM1 and TM2, with deviations of 0.37 in Egoless Individaul Task Assignment (Egoless ITA) and 0.36 in Egoless Group Task Assignment (Egoless GTA), respectively. These deviations are less than half of the average belief biases observed regarding the productivity ratio between TM1 and TM2 (as shown in Figure 8b). Moreover, receiving correct signals does not significantly increase belief biases in the productivity ratio. The causal effect of receiving an incorrect signal in Egoless ITA, 0.11, is not statistically different from zero (F-statistic: 2.24, p = 0.137). Similarly, the causal effect in Egoless GTA, 0.07, is also not significantly different from zero (F-statistic: 1.49, p = 0.225). This is in sharp contrast to the significant effect of receiving incorrect signals on beliefs about the productivity ratio between TM1 and TM2. After controlling for potential impacts of TM1's type, the effect of receiving incorrect signals slightly increases to 0.144 in Egoless ITA (F-statistic: 3.76, p = 0.0553) and 0.104 in Egoless GTA (F-statistic: 3.82, p = 0.054) (See Column (7)-(12) of Table A.8). We also find that the causal effect of receiving incorrect signals about TM1 in Egoless ITA, 0.11, is not statistically different from the causal effect in Egoless GTA, 0.07 (F-statistic: 0.187, p =0.666). This confirms the comparability of belief biases between Egoless ITA and Egoless GTA.

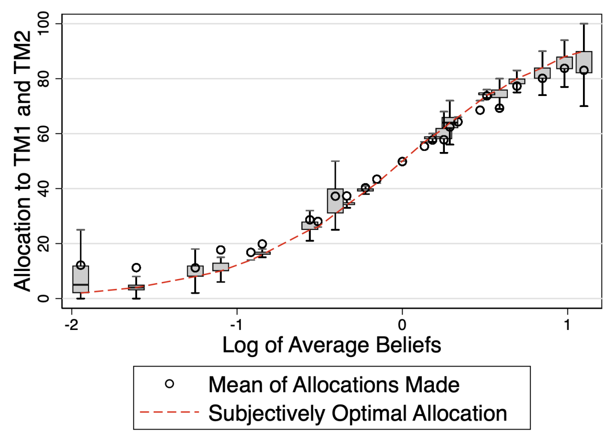
[9] F-statistic: 1.159 (p =0.282 for the difference in difference between *Second-Lowest* and *Average* types; F-statistic: 0.006 (p =0.939) for the difference in difference between *Second-Highest* and *Average* types; F-statistic: 0.485 (p =0.486) for the difference in difference between *Highest* and *Average* types.

4.1.3 Allocation choices

This section examines if subjects behave optimally conditional on beliefs. Figure 10 shows box whisker plots of allocation choices made by subjects. Experiment A and Experiment B are pooled. Figure 10a shows allocation choices to TM1 in Individual Task Assignment (ITA) conditional on beliefs about the productivity ratio between TM1 and TM2. Figure 10b shows allocation choices to a group of TM1 and TM2 in Group Task Assignment (GTA) as a function of their beliefs about the average productivity of TM1 and TM2. In both figures, the mean allocations are overlaid on the box whisker plots and indicated by circles. The red dashed line represents the theoretical benchmark that subjects should choose in order to maximize their perceived expected output based on their beliefs.[10] The figures shows that subjects make allocation choices that are predominantly consistent with subjectively optimal actions.



(a) Individual Task Assignment



(b) Group Task Assignment

Figure 10: Subjectively optimal choices

Note: Panel (a) displays box whisker plots of allocation choices conditional on beliefs about the productivity ratio between TM1 and TM2. Panel (b) displays box whisker plots of allocation choices conditional on beliefs about the average productivity of TM1 and TM2. In both figures, the mean allocations are overlaid on the box whisker plots and indicated by circles. The red dashed line is the benchmark if subjects choose to optimize their allocation based on their beliefs. Experiment A and Experiment B are pooled.

We regress choice allocations on the theoretical benchmark. If subjects subjectively optimize their allocation choices conditional on their beliefs, the coefficient should be 1. Standard errors are clustered at the individual level.

Figure 11 shows the regression results overlaid with the scatter plots of allocation choices relative to the theoretical benchmark. Individual Task Assignment (ITA) and Group Task

[10]Figure B.1 and B.2 provide the figures by Experiment A and Experiment B separately.

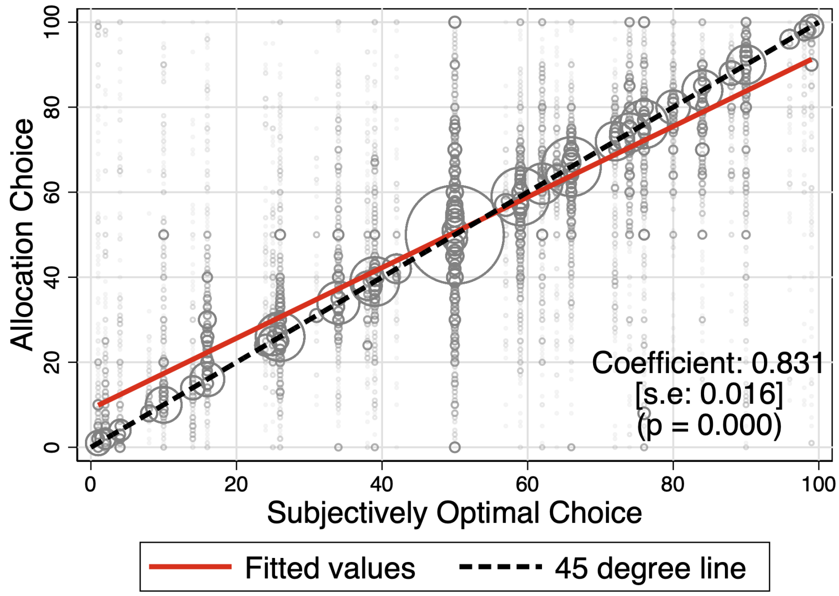


Figure 11: Linear fit of allocation choice on subjectively optimal choice

Note: We regress choice allocations on the theoretical benchmark that subjects should choose in order to maximize their perceived expected output based on their beliefs. The figure shows the regression results overlaid with the scatter plots of allocation choices relative to the theoretical benchmark. Individual Task Assignment (ITA) and Group Task Assignment (GTA) in Experiment A and Experiment B are pooled. The size of the markers indicates the frequency of observations. The red solid line represents the fitted values of the regression estimates, while the black dashed line represents the 45-degree line as the benchmark.

Assignment (GTA) in Experiment A and Experiment B are pooled.[11] The size of the markers indicates the frequency of observations. The red solid line represents the fitted values of the regression estimates, while the black dashed line represents the 45-degree line as the benchmark.

The figure supports that allocation choices made in our experiments are optimal, conditional on subjects' beliefs. For each subjectively optimal choice, we observe the most frequently selected allocation lying on the 45-degree line. This suggests that subjects tend to choose the optimal allocation based on their beliefs. The coefficient of the linear fit is 0.831 (F-statistic: 2607.42, $p = 0.000$).

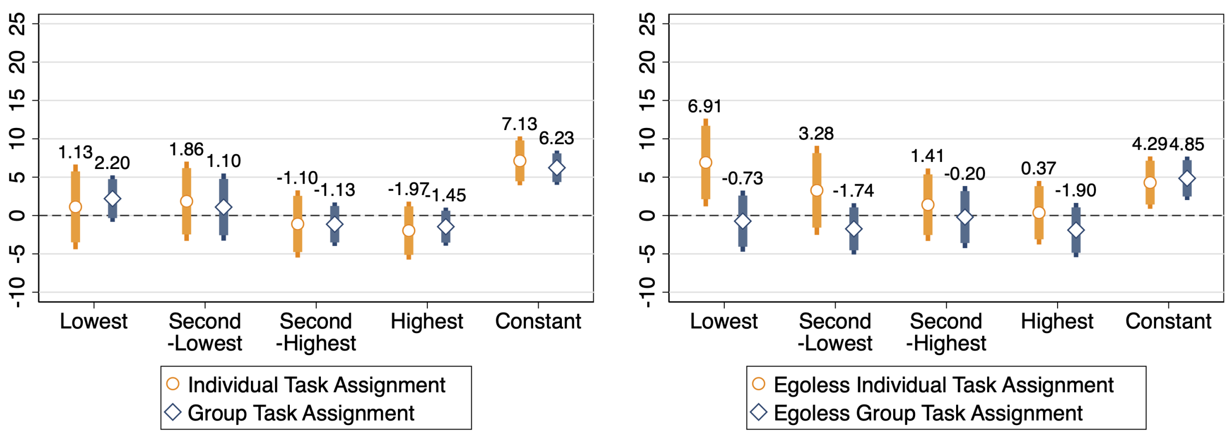
Next, we investigate if there are systematic differences in decision-making across subject types. For example, if being assigned as a *Lowest* or *Second-Lowest* type reflects their lack of attention or engagement in Part 1 of the experiment, it is possible that their decision-making in Part 2 deviates further from their subjectively optimal choice. Moreover, *Lowest* types in Experiment A may tend to allocate more to TM1 (self) due to a desire for control and to

[11] Figures B.3 and B.4 provide the figures by Experiment A and Experiment B separately.

avoid uncertainty regarding TM2's productivity (Benoît et al., 2022).

We define *Abs. Choice Errors* as the absolute difference between the subjectively optimal action and their actual choice. Figure 12 displays the regression coefficients obtained when we regress *Abs. Choice Errors* on the indicators of subject types. The constant captures the average deviation of *Average* types (the omitted category). The coefficients of the indicators for each subject type reveal how much further allocation choices of each type deviate from the subjectively optimum in comparison to *Average* types.

Figure 12 confirms, regardless of subject types, all subjects strive to make optimal decisions, albeit based on their (biased) beliefs. There are no significant differences across subject types. The exception is *Lowest* types in (Egoless) Individual Task Assignment of Experiment B. They deviate from their subjectively optimal choice by 7 hypothetical projects compared to *Average* types (p=0.018).



(a) Experiment1

(b) Experiment B

Figure 12: Absolute errors in choices by subject types

Note: The dependent variable is *Abs. Choice Errors*, defined by the absolute difference between the subjectively optimal action and their actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional choice errors that *Lowest* types make compared to *Average* types. Error bars represent 95% and 90% confidence intervals. In all regressions, we control for rounds, and standard errors are clustered at the individual level. Table A.9 and A.10 report the full regression estimates.

4.2 Treatment effects

Section 4.1.1 establishes that there are no differences in subjects' true productivities and their beliefs about productivities between Individual Task Assignment (ITA) and Group Task Assignment (GTA). In Section 4.1.2, it is shown that subjects in both treatments respond to the incentives imposed by the treatments. Specifically, subjects in ITA update their beliefs about the productivity ratio between TM1 and TM2 more than those in GTA. Conversely, beliefs about the average productivity between TM1 and TM2 are closer to the truth in GTA. Section 4.1.3 confirms that subjects make subjectively optimal actions in both treatment conditions. Consequently, one would expect no difference in efficiency. If there is a difference, it would be reasonable to conjecture that ITA leads to more efficient outcomes because it necessitates learning the productivity of each team member (to determine the productivity ratio between TM1 and TM2) and, consequently, their average.

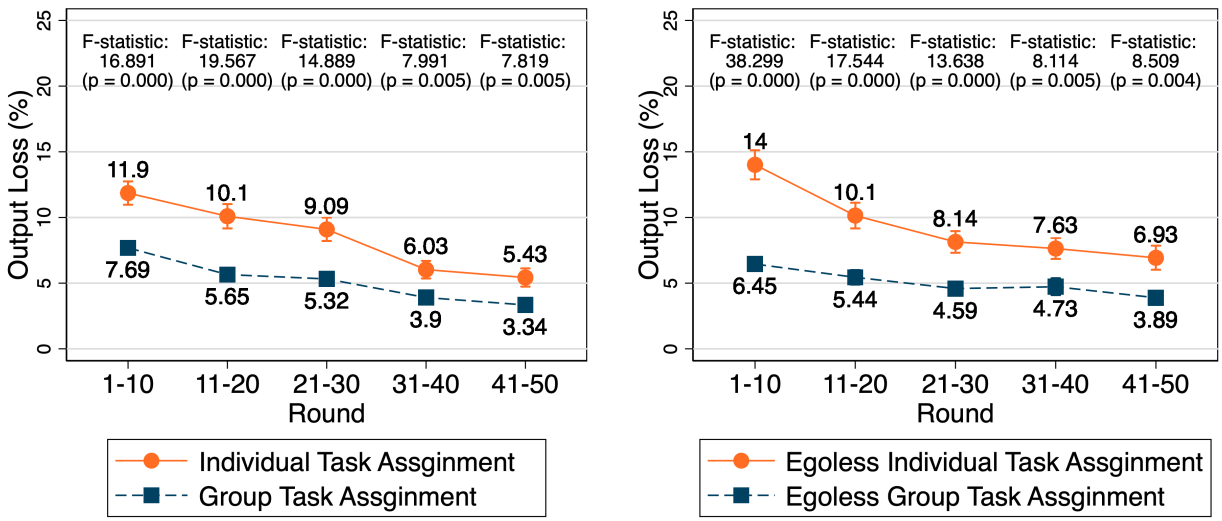
Nonetheless, this section demonstrates that Group Task Assignment (GTA) results in more efficient outcome than Individual Task Assignment (ITA). Allocative efficiency is measured using *Output Loss (%)*, which represents the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). We divide the difference between the highest possible expected output and the actual expected output by the difference between the highest possible expected output and the lowest possible expected output.[12] This measure of allocative efficiency accounts for variations in potential output levels across treatment conditions and across the productivities of TM1 and TM2.

4.2.1 Treatment effects at the aggregate level

Figure 13 presents the average output loss by treatments. Figure 13a shows the results of Experiment A, while Figure 13b displays the results of Experiment B. We regress *Output Loss (%)* on indicators for every ten rounds to calculate the average. The figure displays the predicted values, and the error bars represent clustered standard errors at the individual level. These figures highlight two key findings.

First, output loss decreases over time in both Individual Task Assignment (ITA) and Group Task Assignment (GTA) for both Experiment A and Experiment B. The changes in average output loss from round 1-10 to round 41-50 are statistically significant. In Experiment A, the F-statistic for the difference in ITA is 57.45 (p =0.000), and for GTA, the F-statistic is 82.17 (p =0.000). In Experiment B, the F-statistic for ITA is 46.45 (p =0.000), and for GTA, it is 17.74 (p =0.001). The decreasing pattern implies that subjects are actively engaging in the experiment, making better allocation choices over time.

[12] Using the notations in Section 2, $OutputLoss(\%) = \frac{\max \mu(x|a_1,a_2) - \mu(x|b_1,b_2)}{\max \mu(x^{chosen}|a_1,a_2) - \min \mu(x|a_1,a_2)}$ where x^{chosen} is the allocation choice chosen by a subject.



(a) Experiment A

(b) Experiment B

Figure 13: Aggregate output loss

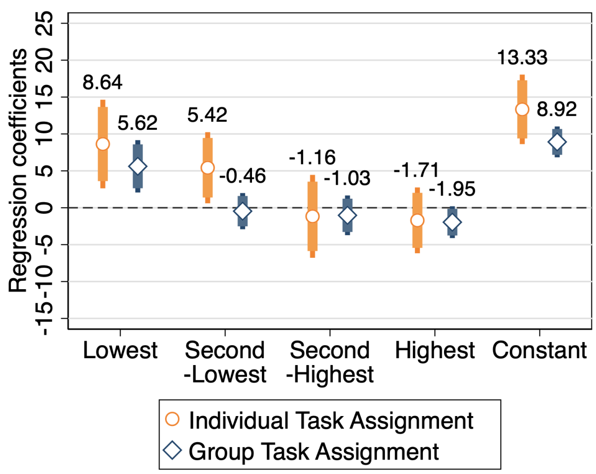
Notes: The figure presents the average output loss over time by treatments. *Output loss (%)* is defined as the proportion of additional potential output that a subject could have achieved if they had made optimal decisions. To calculate the average, we regress output loss on indicators for every ten rounds. The figure displays the predicted values, and the error bars represent clustered standard errors at the individual level. The reported F-statistics and p-values in the figures represent the test for equality of *Output loss (%)* between Individual Task Assignment (ITA) and Group Task Assignment (GTA) for every ten rounds. In Experiment A, the F-statistic for the difference in ITA is 57.45 (p =0.000), and for GTA, the F-statistic is 82.17 (p =0.000). In Experiment B, the F-statistic for ITA is 46.45 (p =0.000), and for GTA, it is 17.74 (p =0.001).

Second, the overall output loss is greater in Individual Task Assignment (ITA) than in Group Task Assignment (GTA). For every tenth round, the output loss of ITA is significantly higher than that of GTA at a 1% significance level. The F-statistics are reported in the figure.

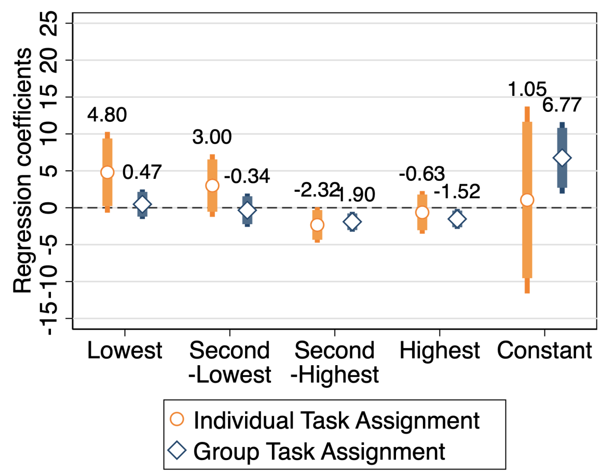
While our measure of allocative efficiency, *Output loss (%)*, is adjusted for differences in achievable output levels, some may not be fully convinced. To address this skepticism, we compare allocative efficiency across subjects within each treatment condition in Experiment A. For instance, *Output loss (%)* for *Lowest* types is expected to be greater than for other subject types in Individual Task Assignment (ITA) because their beliefs are more biased, as shown in Section 4.1.2. If Group Task Assignment (GTA) yields more efficient outcomes in general, regardless of biased beliefs, *Lowest* types in GTA are also expected to experience greater inefficiency as their belief biases are also the largest. However, if we observe no more inefficient outcomes among the *Lowest* types in GTA, we can conclude that it is GTA that suppresses the harmful effect of biased beliefs. Comparisons of *Output loss (%)* between subjects who receive correct and incorrect signals about TM1 in Experiment B allow us to establish the causal effect of ITA and GTA on allocative efficiency.

4.2.2 Treatment effects in Experiment A

Figure 14 shows the differences in output loss across subject types in Experiment A. We regress *Output loss (%)* on the indicators of subject types. The coefficients indicate the differences among subject types compared to *Average* types (the omitted category). For instance, the coefficient of *Lowest* represents the additional percentage points of output loss incurred by *Lowest* types in comparison to Average types. We control for rounds, and standard errors are clustered at the individual level. Table A.11 report the full regression estimates.



(a) Round 1-10



(b) Round 41-50

Figure 14: Treatment effects in Experiment A

Notes: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences among subject types compared to *Average* types (the omitted category). For instance, the coefficient of *Lowest* represents the additional percentage points of output loss incurred by *Lowest* types in comparison to *Average* types. We control for rounds, and standard errors are clustered at the individual level. Table A.11 report the full regression estimates.

Figure 14a presents the results during the first ten rounds. In both Individual Task Assignment (ITA) and Group Task Assignment (GTA), *Lowest* types incur significantly greater inefficiency by 8.64 (p =0.005) and 5.62 percentage points (p =0.002), respectively. Additionally, in ITA, *Second-Lowest* types also experience higher output loss than *Average* types by 5.42 percentage points (p =0.028).

Figure 14b shows the results during the last ten rounds. In Individual Task Assignment (ITA), the efficiency gap between *Lowest* and *Average* types persists. *Lowest* types incur the greatest inefficiency, with their output loss being 4.8 percentage points greater than that of *Average* types ($p = 0.085$). *Second-Lowest* types incur the second greatest inefficiency, although the difference of 3 is not statistically significant at the 10% significance level ($p = 0.163$). In contrast, the efficiency gap between *Lowest* and *Average* types disappears in Group Task Assignment (GTA). The coefficient of *Lowest* types is not statistically different from zero ($p = 0.64$).

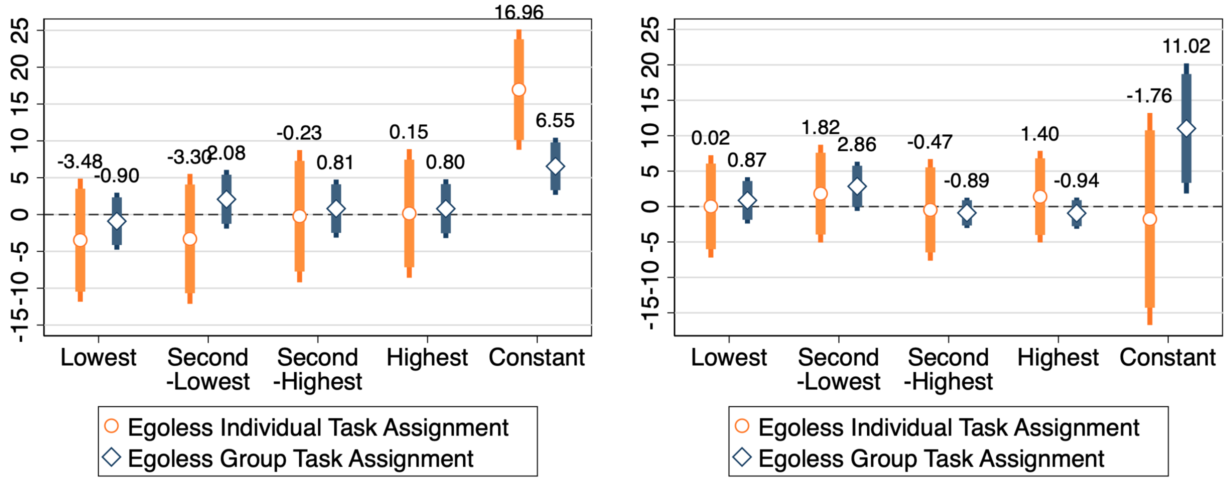
In summary, this section has demonstrated that biased beliefs have a more detrimental impact in Individual Task Assignment (ITA) compared to Group Task Assignment (GTA). To account for potential differences in team output levels between ITA and GTA, we analyze heterogeneity in *Output loss (%)* within each treatment. Finally, we test whether the heterogeneous effects across subject types significantly differ between treatments. To do so, we conduct a regression analysis with the indicators for subject types, treatment indicators, and their interactions. We pool all 50 rounds. The regression controls for rounds. A joint test of the interactions being zero captures the treatment effects while fully controlling for structural differences between ITA and GTA. The result suggests a weak treatment effect. We fail to reject the null hypothesis that the heterogeneous effects are different from ITA and GTA (F-statistic: 1.644, p=0.164). See Table B.1.

4.2.3 Treatment effects in Experiment B

In Experiment A, TM1 represents the subject's self, and thus, the subject's type plays a crucial role. On one hand, subject types represent individuals, that is the subject's self, making allocation decisions. On the other hand, they also represent individuals, specifically TM1, whose productivity determines the objective environment.

On the contrary, in Experiment B, TM1 represents a randomly selected stranger. The productivity of the subject's self is irrelevant to the decision problem subjects are facing. Consequently, we do not expect subject types to correlate with output loss in Experiment B. If such a correlation does emerge, it suggests confounding factors, such as overall inattention during the experiment, affecting the behavior of *Lowest* type subjects in Experiment A.

Furthermore, we anticipate minimal influence from TM1's types on output loss in Experiment B. Unlike subjects in Experiment A, especially those with below-average productivity, subjects in Experiment B are unlikely to consistently maintain biased beliefs about TM1 (e.g., Zell et al., 2020). If we do observe a correlation between TM1's types and output loss in Experiment B, it challenges our findings of Experiment A that greater belief biases among *Lowest* types are associated with greater inefficiency in Individual Task Assignment (ITA); and that Group Task Assignment (GTA) features the protective property against the belief



(a) Round 1-10

(b) Round 41-50

Figure 15: Allocative efficiency across subject types in Experiment B

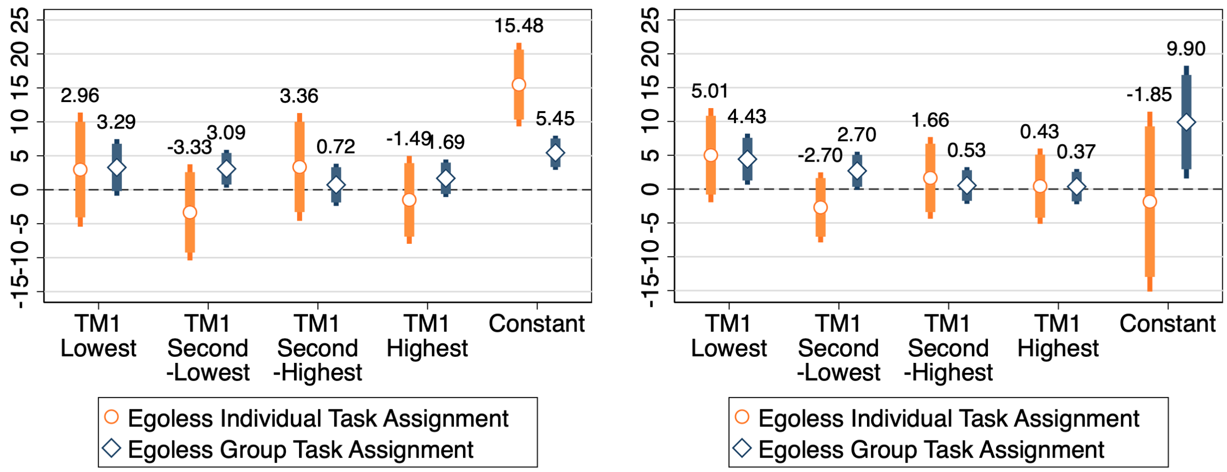
Notes: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences among subject types compared to *Average* types (the omitted category). For instance, the coefficient of *Lowest* represents the additional percentage points of output loss incurred by *Lowest* types in comparison to *Average* types. We control for rounds, and standard errors are clustered at the individual level. Table A.12 reports the full regression estimates.

biases.

We find no supporting evidence for these possibilities. Figure 15 presents an identical figure to Figure 14, using data from Experiment B. Table A.12 report the full regression estimates. Ruling out the first concern, Figure 15 demonstrates that output loss is not correlated with subject types in Experiment B. Figure 16 presents the average output loss by TM1's type in Experiment B. Table A.13 report the full regression estimates. To address the second concern, The figure shows that the efficiency gaps between Individual Task Assignment (ITA) and Group Task Assignment (GTA), shown in Figure 13b, are not attributed to TM1's types.

The impact of signals is of interest in Experiment B. The comparison in allocative efficiency between subjects who receive correct and incorrect signals about TM1 allows us to identify the causal effect of belief biases.

Figure 17 shows the causal effect of receiving incorrect signals about TM1 on allocative efficiency. We regress *Output loss (%)* on the indicator of receiving incorrect signals about TM1. The coefficients captures the difference in output loss between subjects who receive incorrect signals and those who receive correct signals (the omitted category). We control



(a) Round 1-10

(b) Round 41-50

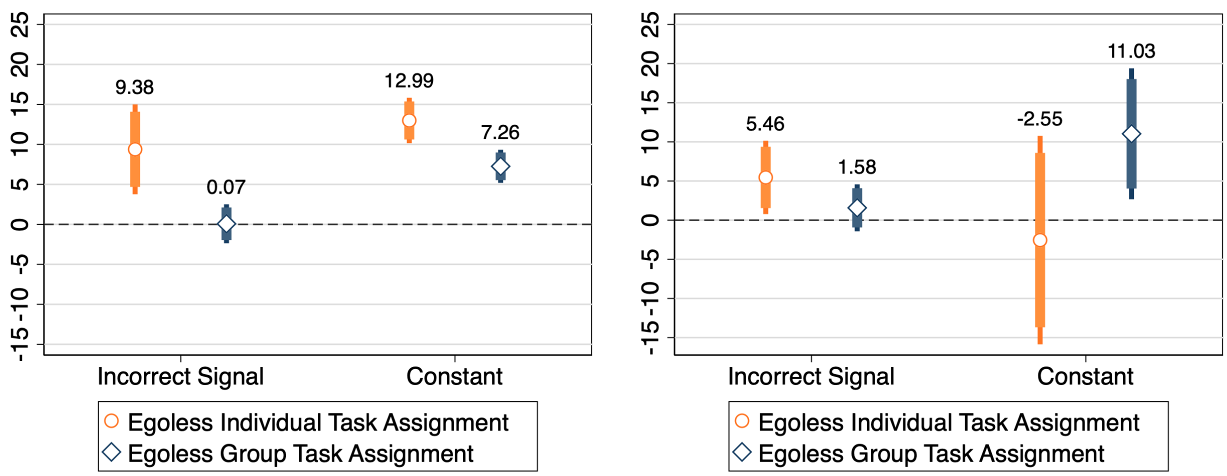
Figure 16: Allocative efficiency across TM1's types in Experiment B

Notes: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *TM1 Lowest*, *TM1 Second-Lowest*, *TM1 Second-Highest*, and *TM1 Highest* refer to the indicators of TM1's types with whom the subject is matched. The coefficients indicate the differences among subjects matched with each type of TM1 compared to those whose TM1 is a *Average* type (the omitted category). For instance, the coefficient of *TM1 Lowest* represents the additional percentage points of output loss incurred by subjects whose TM1 is *Lowest* types in comparison to those whose TM1 is *Average* types. We control for rounds, and standard errors are clustered at the individual level. Table A.13 reports the full regression estimates.

for rounds, and standard errors are clustered at the individual level. Table A.14 report the full regression estimates.

Figure 17a presents the results during the first ten rounds. In Individual Task Assignment (ITA), subjects who receive incorrect signals about TM1 experience a 9.38 percentage points higher output loss compared to those who receive correct signals (p =0.001). In contrast, receiving incorrect signals has an insignificant effect in Group Task Assignment (GTA) (p =0.001). Figure 17b presents the results during the last ten rounds. In ITA, the effect size is attenuated to 5.46, but it remains significant (p =0.023). In GTA, the effect still remains insignificant.

In short, Figures 15 and 16 complement Experiment A by addressing potential concerns. Moreover, Figure 17 reinforces the findings from Experiment A that Group Task Assignment (GTA) provides protection against inefficiency caused by biased beliefs. To draw a conclusion on treatment effects, we conduct a regression analysis with the indicators for receiving incorrect signals, treatment indicators, and their interactions. We pool all 50 rounds. The regression controls for rounds. A test of the interaction term being zero captures the



(a) Round 1-10

(b) Round 41-50

Figure 17: Treatment effects in Experiment B

Notes: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Incorrect Signal* is the indicators of of receiving incorrect signals about TM1. The coefficients indicate the causal effect of receiving incorrect signal about TM1 on output loss. We control for rounds, and standard errors are clustered at the individual level. Table A.14 report the full regression estimates.

treatment effects while fully controlling for structural differences between ITA and GTA. The result indicates strong treatment effects. The effect of receiving incorrect signals significantly differs between treatments (F-statistic: 4.283, p=0.004). See Table B.2.

4.2.4 Mechanisms

So far, we have established that Individual Task Assignment (ITA) and Group Task Assignment (GTA) yield different allocative efficiency outcomes in the presence of equally biased beliefs. Our findings from both Experiment A and Experiment B reinforce each other, supporting the conclusion that GTA provides an environment where biased beliefs have a less harmful effect.

Next, we delve into the mechanisms. First, in Individual Task Assignment (ITA), making the optimal choice necessitates knowledge of the productivity ratio between TM1 and TM2. Therefore, we anticipate that the greater inefficiency observed among the *Lowest* types is primarily driven by their belief biases in the productivity ratio in ITA. Second, in Group Task Assignment (GTA), the optimal choice requires knowledge of the average productivity of TM1 and TM2. Consequently, we expect that errors in beliefs about the average productivity drive a positive output loss in GTA. Third, we anticipate that errors in choosing the optimal allocation choice, conditional on beliefs, have little impact.

To quantify the effects of each of the three factors on allocative efficiency, we regress *Output loss (%)* on three regressors: i) *Abs. Errors in Log Ratio on Beliefs* represent the absolute difference between the log ratio of beliefs about the productivity of TM1 and the beliefs about the productivity of TM2, and the log ratio of the (true) productivity of TM1 and the (true) productivity of TM2. ii) *Abs. Errors in Log Average Beliefs* denote the absolute difference between the log of beliefs about the average productivity of TM1 and TM2, and the log of the (true) average productivity of TM1 and TM2. iii) *Abs. Choice Errors* indicate the absolute difference between the (subjectively) optimal allocation choice and the actual choice.

We use absolute terms because the productivity space is bounded from *Lowest* to *Highest*. For example, errors in beliefs about the productivity of TM1 would always be upwardly biased for those whose TM1 is a *Lowest* type and downwardly biased for those whose TM1 is a *Highest* type. The aim of this specification is to identify whether these potential errors in beliefs and choices have an impact on allocative efficiency and, if so, to what extent.

Table 2 reports the regression estimates for Individual Task Assignment (ITA). The first two columns are from Experiment A, and the third and fourth columns report the estimates in Experiment B. The regressions aggregate data from 50 rounds. We control for rounds, and standard errors are clustered at the individual level are in parentheses.

Column (1) and (3) show that a one-unit increase in *Abs. Errors in Log Ratio on Beliefs* results in an increase in output loss by 13-14.6 percentage points, while the effect of *Abs. Errors in Log Average Beliefs* is less than a fifth of that magnitude (2.03-2.62). *Abs. Choice Errors* has no discernible effect on output loss.

In Column (2), we add the indicators of subject types as regressors. The estimates shows belief biases, particularly errors in beliefs about the productivity ratio between TM1 and TM2, account for the effects associated with subject types observed in Experiment A. As depicted in Figure 14, *Lowest* types experience significantly greater inefficiency compared to *Average* types when potential errors are not controlled for. Pooling all rounds, the output loss of *Lowest* types is greater by 6.36 percentage points (p =0.005), while *Second-Lowest* types encounter greater inefficiency by 3.27 percentage point (p = 0.087). (See Column (3) of Table A.11.) However, these effects vanish after accounting for the biases (the coefficient of *Lowest* is -0.763, p =0.615; the coefficient of *Lowest* is 1.047, p =0.417).

In Column (4), we investigate the extent to which the causal effect of receiving incorrect signals about TM1 is mediated by belief errors. We add the indicator of receiving incorrect signals in the regression. The estimates confirm that errors in beliefs about the productivity ratio between TM1 and TM2 have the most significant impact. Receiving incorrect signals

Table 2: Mechanisms in Individual Task Assignment

	DV: Output Loss (%)			
	Experiment 1		Experiment 2	
	(1)	(2)	(3)	(4)
Abs. Errors in Log Ratio of Beliefs	12.965***	13.032***	14.551***	14.324***
	(1.000)	(1.090)	(0.849)	(0.804)
Abs. Errors in Log of Average Beliefs	2.622***	2.580***	2.031**	1.778**
	(0.836)	(0.944)	(0.794)	(0.800)
Abs. Choice Errors	0.068	0.064	-0.097*	-0.093*
	(0.059)	(0.059)	(0.055)	(0.056)
Lowest		-0.763		
		(1.514)		
Second-Lowest		1.047		
		(1.284)		
Second-Highest		-0.053		
		(1.515)		
Highest		-0.816		
		(1.191)		
Incorrect signal				2.257*
				(1.267)
Round	-0.066***	-0.066***	-0.025	-0.028
	(0.018)	(0.019)	(0.017)	(0.017)
Constant	0.765	0.885	-0.004	-0.324
	(0.792)	(1.075)	(0.661)	(0.661)
Total Observations	5071	5071	4854	4854
Num. of Individuals	102	102	99	99

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Abs. Errors in Log Ratio on Beliefs* represent the absolute difference between the log ratio of beliefs about the productivity of TM1 and the beliefs about the productivity of TM2, and the log ratio of the (true) productivity of TM1 and the (true) productivity of TM2. *Abs. Errors in Log Average Beliefs* denote the absolute difference between the log of beliefs about the average productivity of TM1 and TM2, and the log of the (true) average productivity of TM1 and TM2. *Abs. Choice Errors* indicate the absolute difference between the (subjectively) optimal allocation choice and the actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. *Incorrect Signal* is the indicators of of receiving incorrect signals about TM1. The coefficients indicate the causal effect of receiving incorrect signal about TM1 on output loss. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance.

about TM1 increases output loss by 6.082 percentage points ($p = 0.003$). (See Column (3) of Table A.14.) However, after accounting for belief biases, the causal effect reduces to 2.257 (p =0.078).

Table 3 reports the regression estimates for Group Task Assignment (GTA). The first two columns are from Experiment A, and the third and fourth columns report the estimates in Experiment B. The regressions aggregate data from 50 rounds. We control for rounds, and standard errors are clustered at the individual level are in parentheses.

Table 3 highlights the role of beliefs about the average productivity of TM1 and TM2 in Group Task Assignment (GTA). Columns (1) and (3) indicate that a one-unit increase in *Abs. Errors in Log Average Beliefs* leads to an increase in output loss by 12.5-15.3 percentage points, while *Abs. Errors in Log Ratio on Beliefs* have insignificant impacts. These GTA findings contrast with the results in Individual Task Assignment (ITA), where *Abs. Errors in Log Ratio on Beliefs* rather than *Abs. Errors in Log Average Beliefs* play a significant role. Additionally, a unit increase in *Abs. Choice Errors* results in 0.21-0.33 percentage points higher output loss.

In Column (2), we add the indicators of subject types as regressors. The estimates shows belief biases, particularly errors in beliefs about the average productivity of TM1 and TM2, absorb the effects associated with subject types observed in Experiment A. Pooling all rounds, the output loss of *Lowest* types is greater by 2.64 percentage points (p =0.013). (See Column (6) of Table A.11.) However, the effect becomes statistically indistinguishable from zero after accounting for belief biases (the coefficient is -0.979, $p = 0.226$).

In Column (4), we included the indicator of receiving incorrect signals in the regression for Experiment B. Receiving incorrect signals about TM1 increases output loss by 1.248 percentage points without controlling for potential errors, but the coefficient is not statistically significant at the 10% significance level ($p = 0.291$). (See Column (6) of Table A.14.) After accounting for belief biases, the causal effect drops further to 0.855 ($p = 0.269$).

In summary, our analyses reveal that the differing efficiency outcomes between Individual Task Assignment (ITA) and Group Task Assignment (GTA), despite equally biased beliefs, stem from the distinct beliefs required to make an optimal choice in these treatments. Biased beliefs about the productivity ratio of TM1 and TM2 lead to inefficiency in ITA, whereas biased beliefs about the average productivity of TM1 and TM2 lead to inefficiency in GTA. Because learning the average productivity of TM1 and TM2 is less susceptible to the influence of erroneous beliefs about TM1 compared to learning the productivity ratio of TM1 and TM2, GTA provides protection against the negative consequences of belief biases.

Table 3: Mechanisms in Group Task Assignment

	DV: Output Loss (%)			
	Experiment 1		Experiment 2	
	(1)	(2)	(3)	(4)
Abs. Errors in Log Ratio of Beliefs	0.183	0.331	0.106	-0.023
	(0.228)	(0.259)	(0.312)	(0.329)
Abs. Errors in Log of Average Beliefs	12.526***	12.771***	15.333***	15.228***
	(1.031)	(1.105)	(2.572)	(2.564)
Abs. Choice Errors	0.207***	0.207***	0.327***	0.332***
	(0.064)	(0.064)	(0.089)	(0.089)
Lowest		-0.979		
		(0.807)		
Second-Lowest		-0.363		
		(0.951)		
Second-Highest		-0.131		
		(0.572)		
Highest		0.058		
		(0.597)		
Incorrect signal				0.855
				(0.770)
Round	-0.071***	-0.070***	-0.026	-0.026
	(0.011)	(0.012)	(0.018)	(0.018)
Constant	1.191*	1.206	-0.580	-0.636
	(0.613)	(0.851)	(0.934)	(0.922)
Total Observations	10873	10873	4627	4627
Num. of Individuals	218	218	93	93

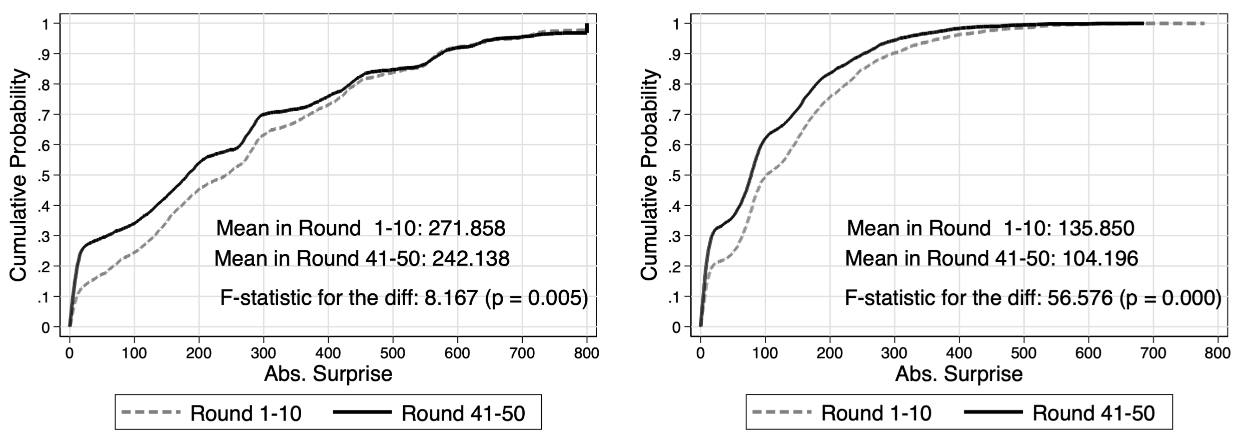
Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Abs. Errors in Log Ratio on Beliefs* represent the absolute difference between the log ratio of beliefs about the productivity of TM1 and the beliefs about the productivity of TM2, and the log ratio of the (true) productivity of TM1 and the (true) productivity of TM2. *Abs. Errors in Log Average Beliefs* denote the absolute difference between the log of beliefs about the average productivity of TM1 and TM2, and the log of the (true) average productivity of TM1 and TM2. *Abs. Choice Errors* indicate the absolute difference between the (subjectively) optimal allocation choice and the actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. *Incorrect Signal* is the indicators of of receiving incorrect signals about TM1. The coefficients indicate the causal effect of receiving incorrect signal about TM1 on output loss. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance.

4.3 Stable beliefs

Section 4.3 examines if subjects' beliefs satisfy the sufficient condition: expectation matches reality if beliefs are stable.

4.3.1 Expectation-reality gaps

Figure 18 displays the cumulative distributions of *Abs. Surprise*. Figure 18a presents the distribution for Individual Task Assignment (ITA), and Figure 18b presents the distribution for Group Task Assignment (GTA). We pool Experiment A and Experiment B. Values exceeding 800 are capped at 800.



(a) Individual Task Assignment

(b) Group Task Assignment

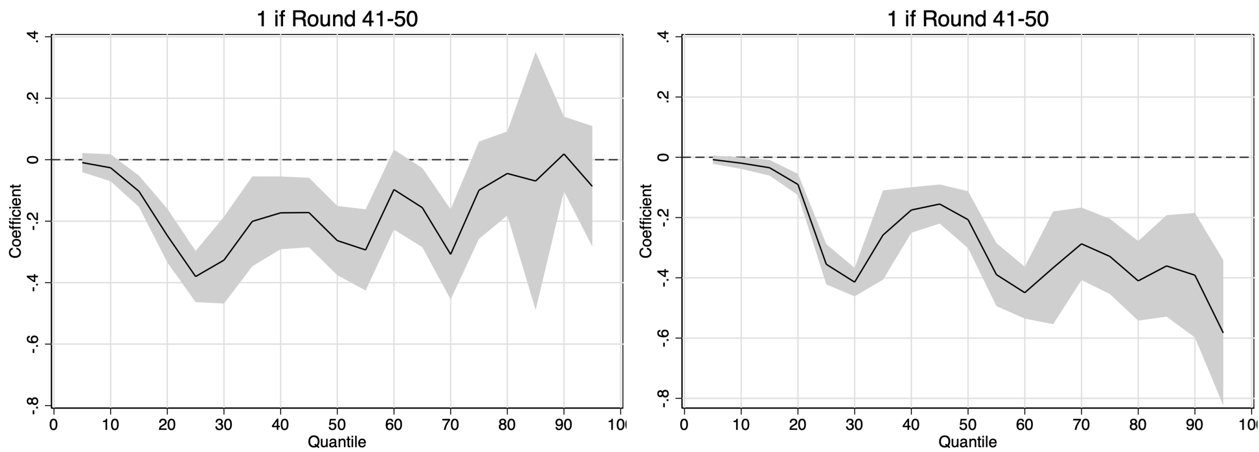
Figure 18: Cumulative distributions of *Abs. Surprise*

Notes: The figure presents the cumulative distributions of *Abs. Surprise*, which is the absolute difference between the perceived expected output and the actual expected output. We pool Experiment A and Experiment B. Values exceeding 800 are capped at 800.

In both treatments, the distributions shift to the left over time. The average *Abs. Surprise* in Individual Task Assignment (ITA) during the first 10 rounds is 271.9, and it decreases by 29.72. The F-statistic for the test for the equality between the first and last ten rounds is 8.17 (p =0.005). The average *Abs. Surprise* in Group Task Assignment (GTA) during the first 10 rounds is 135.85, and it decreases by 31.65. The F-statistic for the test for the equality between the first and last ten rounds is 56.58 (p =0.000).

However, in both cases, *Abs. Surprise* is significantly larger than zero in the last ten rounds (F-statistic: 282.46, p=0.000 in ITA and F-statistic: 527.19, p=0.000). All standard errors are clustered at the individual level.

To examine whether subjects who experience more significant gaps between their expectations and reality adjust their beliefs more rapidly than others, we run quantile regressions. The dependent variable is *Abs. Surprise*. To facilitate comparison between Individual Task Assignment (ITA) and Group Task Assignment (GTA), we standardize the variable within each treatment, resulting in a mean of zero and a variance of 1. We regress the outcome variable on the indicators for every ten rounds. Standard errors are clustered at the individual level.



(a) Individual Task Assignment

(b) Group Task Assignment

Figure 19: Changes in *Abs. Surprise* by quantiles

Notes: The figure presents the estimates from quantile regressions. The dependent variable is *Abs. Surprise*, which is the absolute difference between the perceived expected output and the actual expected output. To facilitate comparison between Individual Task Assignment (ITA) and Group Task Assignment (GTA), we standardize the variable within each treatment, resulting in a mean of zero and a variance of 1. We regress the outcome variable on the indicators for every ten rounds. Standard errors are clustered at the individual level. We pool Experiment A and Experiment B.

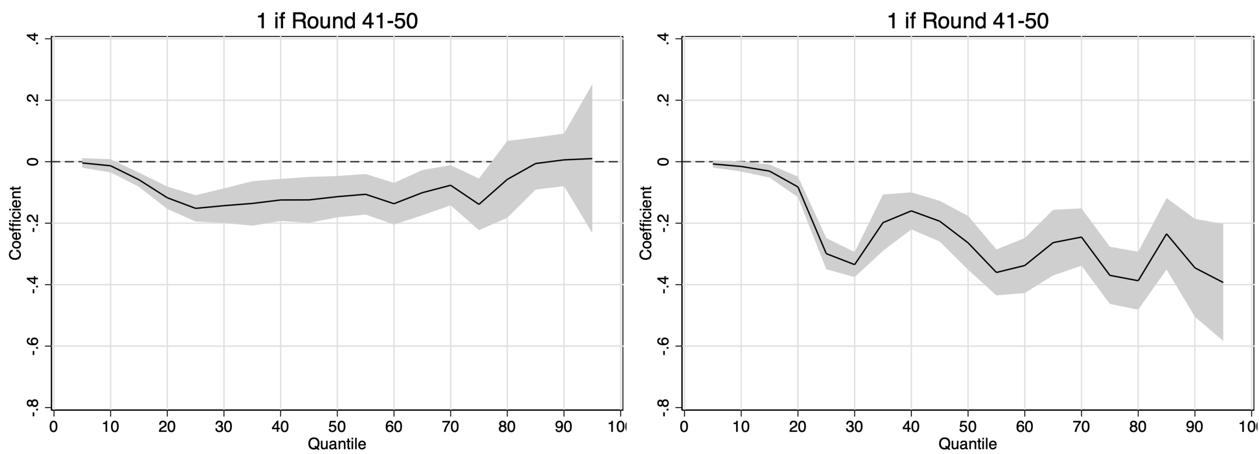
Figure 19 displays the coefficients of the indicator for the last ten rounds (rounds 41-50). Figure 19a presents the results for Individual Task Assignment (ITA), and Figure 19b presents the results for Group Task Assignment (GTA).

These coefficients indicate the changes in *Abs. Surprise* between the first and last ten rounds at every five percentile point. Figure 19a shows that subjects in the 25$^{\text{th}}$ percentile during rounds 1-10 experience the most significant reduction in the expectation-reality gap. Contrary to Hypothesis 3.2, we do not find evidence that subjects with higher initial values of *Abs. Surprise* exhibit a faster decline in this metric. For instance, subjects who encounter *Abs. Surprise* in the above 90th percentile do not experience a significant decrease in the gap.

In contrast, we find evidence that subjects with higher initial values of *Abs. Surprise* experience a faster decline in Group Task Assignment (GTA). Figure 19b shows that, as subjects who initially experience a greater expectation-reality gap in earlier rounds, their gaps decrease more significantly.

4.3.2 Perceived Expectation-reality gaps

The failure to find evidence for Hypothesis 3.1 may be attributed to the fact that subjects perceive and interpret expectation-reality gaps through the lens of their understanding of the environment, i.e., their mental models. To account for this possibility, we regress *Abs. Surprise rel. Expectation* on the indicators for every ten rounds. We standardize the dependent variable within each treatment, resulting in a mean of zero and a variance of 1. Standard errors are clustered at the individual level.



(a) Individual Task Assignment

(b) Group Task Assignment

Figure 20: Changes in *Abs. Surprise rel. Expectation* by quantiles

Notes: The figure presents the estimates from quantile regressions. The dependent variable is *Abs. Surprise rel. Expectation*, which is the absolute difference between the perceived expected output and the actual expected output relative to the perceived expected output. To facilitate comparison between Individual Task Assignment (ITA) and Group Task Assignment (GTA), we standardize the variable within each treatment, resulting in a mean of zero and a variance of 1. We regress the outcome variable on the indicators for every ten rounds. Standard errors are clustered at the individual level. We pool Experiment A and Experiment B.

Figure 20 displays the coefficients of the indicator for the last ten rounds (rounds 41-50) for quantile regressions. Figure 20a presents the results for Individual Task Assignment (ITA), and Figure 20b presents the results for Group Task Assignment (GTA).

The results are qualitatively identical to those in Figure 19. In Individual Task Assignment (ITA), as shown in Figure 20a, large values of *Abs. Surprise rel. Expectation* do not result in a higher reduction in expectation-reality gaps. On the contrary, Figure 20b shows that subjects who initially experience a greater expectation-reality gap in earlier rounds reduce their gaps decrease more significantly in Group Task Assignment (GTA).

5 Conclusion

This paper has identified a condition under which biased beliefs lead to optimal action. We contribute to the literature by introducing a relatively new research agenda to understand when biased beliefs have greater consequences and when they are harmless, rather than focusing on correcting these beliefs. We employ a teamwork setting as a specific context for this study. However, we anticipate that the implications of our findings extend far beyond the realm of teamwork, encompassing other scenarios where individuals must navigate decision-making in the presence of their own biased beliefs.

References

Ayres, Ian, Mahzarin Banaji, and Christine Jolls (2015), "Race effects on ebay." *The RAND Journal of Economics*, 46, 891–917.

Benoît, Jean-Pierre, Juan Dubra, and Giorgia Romagnoli (2022), "Belief elicitation when more than money matters: Controlling for "control"." *American Economic Journal: Microeconomics*, 14, 837–88.

Berg, Joyce E., Lane A. Daley, John W. Dickhaut, and John R. O'Brien (1986), "Controlling preferences for lotteries on units of experimental exchange." *The Quarterly Journal of Economics*, 101, 281–306.

Bodenhausen, Galen V. (1990), "Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination." *Psychological Science*, 1, 319–322.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2012), "Salience theory of choice under risk." *The Quarterly Journal of Economics*, 127, 1243–1285.

Bursztyn, Leonardo, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth (2023), "Justifying Dissent." *The Quarterly Journal of Economics*, 138, 1403–1451.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott (2020), "Misperceived social norms: Women working outside the home in saudi arabia." *American Economic Review*, 110, 2997–3029.

Bursztyn, Leonardo and David Y. Yang (2022), "Misperceptions about others." *Annual Review of Economics*, 14, 425–452.

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016), "otree—an open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 9, 88–97.

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

Clance, Pauline Rose and Suzanne Ament Imes (1978), "The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention." *Psychotherapy*, 15, 241–247.

Coutts, Alexander, Leonie Gerhards, and Zahra Murad (2020), "Who to blame? self-serving attribution bias with multi-dimensional uncertainty."

Edelman, Benjamin and Michael Luca (2014), "Digital discrimination: The case of airbnb.com." *SSRN Electronic Journal*.

Enke, Benjamin (2020), "What You See Is All There Is." *The Quarterly Journal of Economics*, 135, 1363–1398.

Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2023), "Confidence, self-selection, and bias in the aggregate." *American Economic Review*, 113, 1933–66.

Enke, Benjamin and Florian Zimmermann (2019), "Correlation Neglect in Belief Formation."

The Review of Economic Studies, 86, 313–332.

Epley, Nicholas and Thomas Gilovich (2016), "The mechanics of motivated reasoning." *Journal of Economic Perspectives*, 30, 133–40.

Esponda, Ignacio and Demian Pouzo (2016), "Berk–nash equilibrium: A framework for modeling agents with misspecified models." *Econometrica*, 84, 1093–1130.

Esponda, Ignacio, Emanuel Vespa, and Sevgi Yuksel (2022), "Mental models and learning: The case of base-rate neglect."

Exley, Christine L. and Kirby Nielsen (2022), "The gender gap in confidence: Expected but not accounted for." *Working Paper*.

Fiske, Susan T. (2002), "What we know now about bias and intergroup conflict, the problem of the century." *Current Directions in Psychological Science*, 11, 123 – 128.

Fudenberg, Drew and Giacomo Lanzani (forthcoming), "Which Misspecifications Persist?" *Theoretical Economics*.

Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack (2021), "Limit Points of Endogenous Misspecified Learning." *Econometrica*, 89, 1065–1098.

Gagnon-Barstch, Tristan and Benjamin Bushong (2023), "Heterogeneous tastes and social (mis)learning." *Working Paper*.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein (2021), "Channeled attention and stable errors."

Gill, David and Victoria Prowse (2012), "A structural analysis of disappointment aversion in a real effort competition." *American Economic Review*, 102, 469–503.

Gill, David and Victoria Prowse (2019), "Measuring costly effort using the slider task." *Journal of Behavioral and Experimental Finance*, 21, 1–9.

Goette, Lorenz and Marta Kozakiewicz (2022), "Experimental evidence on misguided learning." *Working Paper*.

Golman, Russell, David Hagmann, and George Loewenstein (2017), "Information avoidance." *Journal of Economic Literature*, 55, 96–135.

Graeber, Thomas (Forthcoming), "Inattentive inference." *Journal of the European Economic Association*.

Greiner, Ben (2015), "Subject pool recruitment procedures: Organizing experiments with orsee." *Journal of the Economic Science Association*, 1, 114–125.

Haaland, Ingar K., Christopher Roth, and Johannes Wohlfart (2020), "Designing information provision experiments." CESifo Working Paper 8406, Munich.

Handel, Benjamin and Joshua Schwartzstein (2018), "Frictions or mental gaps: What's behind the information we (don't) use and when do we care?" *Journal of Economic Perspectives*, 32, 155–78.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein (2014), "Learning through noticing: Theory and evidence from a field experiment." *Quarterly Journal of Economics*, 129, 1311–1353.

Heider, Fritz (1944), "Social perception and phenomenal causality." *Psychological Review*, 51, 358–374.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack (2018), "Unrealistic expectations and misguided learning." *Econometrica*, 86, 1159–1214.

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

Hewstone, Miles (1990), "The 'ultimate attribution error'? a review of the literature on intergroup causal attribution." *European Journal of Social Psychology*, 20, 311–335.

Hilton, James L. and William von Hippel (1996), "Stereotypes." *Annual Review of Psychology*, 47, 237–271. PMID: 15012482.

Hoffman, Mitchell and Stephen V. Burks (2020), "Worker overconfidence: Field evidence and implications for employee turnover and firm profits." *Quantitative Economics*, 11, 315–348.

Kendall, Chad and Ryan Oprea (forthcoming), "On the complexity of forming mental models." *Quantitative Economics*.

Kinsler, Josh and Ronni Pavan (2021), "Local distortions in parental beliefs over child skill." *Journal of Political Economy*, 129, 81–100.

Koellinger, Philipp, Maria Minniti, and Christian Schade (2007), ""i think i can, i think i can": Overconfidence and entrepreneurial behavior." *Journal of Economic Psychology*, 28, 502–527.

Lang, Kevin and Ariella Kahn-Lang Spitzer (2020), "Race discrimination: An economic perspective." *The Journal of Economic Perspectives*, 34, 68–89.

Macrae, C. and Galen Bodenhausen (2000), "Social cognition: Thinking categorically about others." *Annual review of psychology*, 51, 93–120.

Malmendier, Ulrike and Geoffrey Tate (2005), "Ceo overconfidence and corporate investment." *The Journal of Finance*, 60, 2661–2700.

Malmendier, Ulrike and Geoffrey Tate (2008), "Who makes acquisitions? ceo overconfidence and the market's reaction." *Journal of Financial Economics*, 89, 20–43.

Marray, Kieran, Nikhil Krishna, and Jarel Tang (2020), "Who to blame? self-serving attribution bias with multi-dimensional uncertainty."

Moore, Don A. and Paul J. Healy (2008), "The trouble with overconfidence." *Psychological review*, 115 2, 502–17.

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa (2021), "Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence, and bias." *American Economic Review*, 111, 324–63.

Niederle, Muriel and Lise Vesterlund (2007), "Do women shy away from competition? do men compete too much?" *The Quarterly Journal of Economics*, 122, 1067–1101.

Rabin, Matthew and Joel L. Schrag (1999), "First impressions matter: A model of confirmatory bias." *The Quarterly Journal of Economics*, 114, 37–82.

Recalde, María P. and Lise Vesterlund (2022), "Gender Differences in Negotiation and Policy for Equalizing Outcomes." In *Bargaining* (Emin Karagözoğlu and Kyle B. Hyndman, eds.), Springer Books, chapter 0, 455–475, Springer.

Santos-Pinto, Luis and Leonidas Enrique de la Rosa (2020), *Overconfidence in Labor Markets*, 1–42. Springer International Publishing, Cham.

Sims, Christopher A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Stinebrickner, Ralph and Todd R. Stinebrickner (2014), "A major in science? initial beliefs and final outcomes for college major and dropout." *The Review of Economic Studies*, 81, 426–472.

Svenson, Ola (1981), "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, 47, 143–148.

Thaler, Richard, C. Sunstein, and John Balz (2012), *Choice Architecture*.

Zell, Ethan, Jason Strickhouser, Constantine Sedikides, and Mark Alicke (2020), "The better-than-average effect in comparative self- evaluation: A comprehensive review and meta-analysis." *Psychological Bulletin*, 146.

Zimmermann, Florian (2020), "The dynamics of motivated beliefs." *American Economic Review*, 110, 337–61.

Proof of Proposition 1

Let $\mathcal{A} = \{10, 30, 50, 70, 90\}$ denote the parameter space. Let $p_1 : \mathcal{A} \to [0, 1]$ denote the belief distribution about $a_1 \in \mathcal{A}$. Let $p_2 : \mathcal{A} \to [0, 1]$ denote the belief distribution about $a_2 \in \mathcal{A}$. Note that there is always a unique (myopically) optimal action $x^* = \arg\max \mu^i(x|a_1, a_2) \in [0, 100]$ for $i \in \{ITA, GTA\}$, and that $\max \sum_{\mathcal{A}} \sum_{\mathcal{A}} p_1(\tilde{a_1}) p_2(\tilde{a_2}) \mu^i(x|\tilde{a_1}, \tilde{a_2}) = \max \mu^i(x|b_1, b_2)$. Therefore, by definition, the following condition is satisfied in equilibrium:

$$\mu^i(x^e|a_1, a_2) = \mu^i(x^e|b_1, b_2) \quad \text{such that} \quad x^e = \arg\max \mu^i(x|b_1, b_2)$$

- Individual Task Assignment:

  Rearranging the equation gives

$$x^e = \frac{100 b_1^2}{b_1^2 + b_2^2} \quad \text{and} \quad (b_1 - \frac{a_1}{2})^2 + (b_2 - \frac{a_2}{2})^2 = \frac{a_1^2 + a_2^2}{4}$$

  The equilibrium action x^e is optimal when it is equal to $x^* = \arg\max \mu^{ITA}(x|a_1, a_2)$. Solving the equation yields $\frac{a_1}{a_2} = \frac{b_1}{b_2}$.

- Group Task Assignment:

  Rearranging the equation gives

$$x^e = \frac{100(b_1 + b_2)^2}{(b_1 + b_2)^2 + 10,000} \quad \text{and} \quad a_1 + a_2 = b_1 + b_2$$

  The equilibrium action x^e is optimal when it is equal to $x^* = \arg\max \mu^{GTA}(x|a_1, a_2)$. Solving the equation yields $a_1 + a_2 = b_1 + b_2$.

Pairs of stable beliefs and stable actions

The table presents all possible pairs of stable, and biased, beliefs in Individual Task Assignment.

$(a_1,$	$a_2)$	$(b_1,$	$b_2)$	$x^*(b_1, b_2)$
10	50	30	30	50
10	70	30	10	90
10	90	50	50	50
30	70	50	50	50
50	10	30	30	50
50	90	70	70	50
70	10	10	30	10
70	30	50	50	50
70	90	90	30	90
90	10	50	50	50
90	50	70	70	50
90	70	30	90	10

A Regression tables

Section 4.1.2: Learning the productivity ratio

Table A.1: Learning the productivity ratio of TM1 and TM2 in Experiment A

	DV: Log Ratio of Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
True Log Ratio (α_1)	0.208***	0.011	0.563***	0.039*	0.416***	0.037**
	(0.046)	(0.015)	(0.070)	(0.023)	(0.051)	(0.016)
Round	-0.001	0.002	0.014***	0.009**	-0.001	0.002***
	(0.008)	(0.004)	(0.005)	(0.004)	(0.001)	(0.001)
Constant	0.071	-0.010	-0.605**	-0.351*	0.070	-0.011
	(0.053)	(0.027)	(0.242)	(0.194)	(0.043)	(0.019)
$\boldsymbol{H_0 : \alpha_1^{ITA} = \alpha_1^{GTA}}$						
F-statistic		16.646		50.299		50.551
p-value		0.000		0.000		0.000
R^2	0.079	0.001	0.437	0.008	0.250	0.006
Total Observations	1019	2178	1000	2178	5071	10873
Num. of Individuals	102	218	102	218	102	218

Note: The table reports the regression estimates of Equation (3) in Experiment A. The dependent variable is *Log Ratio of Beliefs*: the log of the ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2. The independent variable is *True Log Ratio*: the log of the true ratio of the productivity of TM1 to the productivity of TM2. If subjects have the correct belief about the productivity ratio of TM1 and TM2, the coefficient should be 1. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between ITA and GTA. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 6a.

Table A.2: Learning the productivity ratio of TM1 and TM2 in Experiment B

	DV: Log Ratio of Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
True Log Ratio (α_1)	0.303***	0.053	0.637***	0.072	0.529***	0.080
	(0.060)	(0.060)	(0.057)	(0.058)	(0.055)	(0.063)
Round	-0.004	-0.012	-0.002	0.010	0.002	0.001
	(0.010)	(0.009)	(0.008)	(0.007)	(0.002)	(0.001)
Constant	-0.164*	-0.104	-0.028	-0.602*	-0.161**	-0.187***
	(0.086)	(0.075)	(0.355)	(0.304)	(0.066)	(0.064)
$\boldsymbol{H_0 : \alpha_1^{ITA} = \alpha_1^{GTA}}$						
F-statistic		8.815		48.246		29.286
p-value		0.003		0.000		0.000
R^2	0.123	0.006	0.484	0.010	0.338	0.009
Total Observations	980	917	966	929	4854	4627
Num. of Individuals	99	92	97	93	99	93

Note: The table reports the regression estimates of Equation (3) in Experiment B. The dependent variable is *Log Ratio of Beliefs*: the log of the ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2. The independent variable is *True Log Ratio*: the log of the true ratio of the productivity of TM1 to the productivity of TM2. If subjects have the correct belief about the productivity ratio of TM1 and TM2, the coefficient should be 1. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between ITA and GTA. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 6b.

Section 4.1.2: Learning the average productivity

Table A.3: Learning the log of the average productivity in Experiment A

	DV: Log of Average Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Log of True Average (α_1)	0.153***	0.240***	0.149*	0.563***	0.127**	0.460***
	(0.053)	(0.041)	(0.079)	(0.079)	(0.057)	(0.055)
Round	0.001	-0.010**	-0.000	-0.002	0.000	-0.001*
	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)	(0.000)
Constant	3.430***	3.129***	3.454***	1.876***	3.508***	2.217***
	(0.205)	(0.161)	(0.331)	(0.341)	(0.225)	(0.215)
$\boldsymbol{H_0: \alpha_1^{ITA} = \alpha_1^{GTA}}$						
F-statistic		1.685		13.625		17.718
p-value		0.195		0.000		0.000
R^2	0.041	0.080	0.041	0.333	0.026	0.235
Total Observations	1019	2178	1000	2178	5071	10873
Num. of Individuals	102	218	102	218	102	218

Note: The table reports the regression estimates of Equation (3) in Experiment A. The dependent variable is *Log of Average Beliefs*: the log of beliefs about the average productivity of TM1 and TM2. The independent variable is *Log of True Average*: the log of the true average productivity of TM1 and TM2. If subjects have the correct belief about the productivity ratio of TM1 and TM2, the coefficient should be 1. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between ITA and GTA. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 7a.

Table A.4: Learning the log of the average productivity in Experiment B

	DV: Log of Average Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Log of True Average (α_1)	0.136**	0.248***	0.244**	0.481***	0.208***	0.395***
	(0.055)	(0.061)	(0.099)	(0.083)	(0.069)	(0.072)
Round	-0.001	-0.001	0.003	-0.001	-0.000	0.000
	(0.004)	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)
Constant	3.364***	2.953***	2.777***	2.119***	3.077***	2.379***
	(0.219)	(0.245)	(0.448)	(0.342)	(0.273)	(0.286)
$\boldsymbol{H_0: \alpha_1^{ITA} = \alpha_1^{GTA}}$						
F-statistic		1.883		3.347		3.559
p-value		0.172		0.069		0.061
R^2	0.039	0.101	0.084	0.295	0.068	0.213
Total Observations	980	917	966	929	4854	4627
Num. of Individuals	99	92	97	93	99	93

Note: The table reports the regression estimates of Equation (3) in Experiment B. The dependent variable is *Log of Average Beliefs*: the log of beliefs about the average productivity of TM1 and TM2. The independent variable is *Log of True Average*: the log of the true average productivity of TM1 and TM2. If subjects have the correct belief about the productivity ratio of TM1 and TM2, the coefficient should be 1. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. The reported F-statistics and p-values in the figures represent the test for equality of the estimated coefficients between ITA and GTA. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 7b.

Section 4.1.2: Who has more biased beliefs?

Table A.5: Biases in beliefs about the productivity ratio in Experiment A

	DV: Abs. Errors in Log Ratio in Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Lowest	0.719***	0.765***	0.362**	0.801***	0.474***	0.779***
	(0.171)	(0.133)	(0.167)	(0.150)	(0.150)	(0.132)
Second-Lowest	0.263*	0.069	0.101	-0.002	0.138	0.045
	(0.135)	(0.103)	(0.130)	(0.112)	(0.125)	(0.100)
Second-Highest	-0.114	0.011	-0.092	-0.010	-0.094	0.010
	(0.146)	(0.127)	(0.137)	(0.131)	(0.130)	(0.120)
Highest	0.026	0.147	0.022	0.031	0.002	0.095
	(0.141)	(0.130)	(0.126)	(0.126)	(0.120)	(0.122)
Round	-0.017***	0.007*	-0.000	-0.001	-0.007***	-0.001**
	(0.006)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)
Constant	0.712***	0.722***	0.416*	0.823***	0.718***	0.792***
	(0.119)	(0.083)	(0.214)	(0.209)	(0.097)	(0.077)
H_0 : Lowest$_{ITA}$ = Lowest$_{GTA}$						
F-statistic		0.046		3.853		2.338
p-value		0.831		0.051		0.127
H_0 : Second-Lowest$_{ITA}$ = Second-Lowest$_{GTA}$						
F-statistic		1.310		0.362		0.334
p-value		0.253		0.548		0.564
H_0 : Second-Highest$_{ITA}$ = Second-Highest$_{GTA}$						
F-statistic		0.420		0.187		0.343
p-value		0.518		0.666		0.559
H_0 : Highest$_{ITA}$ = Highest$_{GTA}$						
F-statistic		0.403		0.002		0.292
p-value		0.526		0.961		0.590
R^2	0.152	0.121	0.055	0.145	0.097	0.127
Total Observations	1019	2178	1000	2178	5071	10873
Num. of Individuals	102	218	102	218	102	218

Note: The dependent variable is the absolute distance between *Log Ratio of Beliefs*, defined by the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2, and the true productivity ratio of TM1 and TM2. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional belief biases with respect to the productivity ratio of TM1 and TM2 that *Lowest* types make compared to *Average* types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 8a.

Table A.6: Biases in beliefs about the productivity ratio in Experiment B

	DV: Abs. Errors in Log Ratio in Beliefs											
	Round 1-10		Round 41-50		Round 1-50		Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA	(7) ITA	(8) GTA	(9) ITA	(10) GTA	(11) ITA	(12) GTA
Incorrect Signal	0.374***	0.660***	0.225*	0.532***	0.263**	0.584***	0.359***	0.746***	0.240**	0.621***	0.264**	0.678***
	(0.138)	(0.154)	(0.129)	(0.179)	(0.113)	(0.169)	(0.136)	(0.146)	(0.116)	(0.180)	(0.104)	(0.163)
Round	-0.003	0.013	-0.002	-0.007	-0.010***	-0.001	-0.003	0.013*	-0.002	-0.007	-0.010***	-0.001
	(0.008)	(0.008)	(0.007)	(0.006)	(0.001)	(0.001)	(0.008)	(0.008)	(0.007)	(0.006)	(0.001)	(0.001)
TM1 Lowest							0.349*	0.650***	0.599***	0.657***	0.465***	0.649***
							(0.189)	(0.207)	(0.166)	(0.220)	(0.144)	(0.201)
TM1 Second-Lowest							-0.170	0.244	0.062	0.261	-0.049	0.340**
							(0.122)	(0.183)	(0.145)	(0.185)	(0.104)	(0.160)
TM1 Second-Highest							0.026	0.265	0.166	0.235	0.085	0.279
							(0.122)	(0.204)	(0.119)	(0.226)	(0.083)	(0.203)
TM1 Highest							0.125	-0.022	0.180	-0.041	0.124	-0.017
							(0.148)	(0.163)	(0.119)	(0.178)	(0.097)	(0.151)
Constant	0.861***	0.770***	0.571*	1.102***	0.863***	0.866***	0.796***	0.551***	0.353	0.888***	0.733***	0.620***
	(0.075)	(0.072)	(0.330)	(0.271)	(0.060)	(0.065)	(0.095)	(0.143)	(0.344)	(0.295)	(0.064)	(0.121)
H_0 : Incorrect Signal$_{ITA}$ = Incorrect Signal$_{GTA}$												
F-statistic		1.930		1.947		2.518		3.780		3.180		4.602
p-value		0.166		0.165		0.114		0.053		0.076		0.033
R^2	0.043	0.123	0.023	0.080	0.063	0.090	0.089	0.200	0.115	0.162	0.123	0.166
Total Observations	980	917	966	929	4854	4627	980	917	966	929	4854	4627
Num. of Individuals	99	92	97	93	99	93	99	92	97	93	99	93

Note: The dependent variable is the absolute distance between *Log Ratio of Beliefs*, defined by the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2, and the true productivity ratio of TM1 and TM2. *Incorrect Signal* is the indicator of subjects who receive incorrect signals about TM1. The coefficient indicate the causal effect of receiving incorrect signals about TM1 on belief biases. The coefficient of *Incorrect Signal* captures the additional belief biases resulting from receiving incorrect signals. The coefficient of *Incorrect Signal* in Column (1)-(6) captures the additional belief biases resulting from receiving incorrect signals. The coefficient of *Incorrect Signal* in Column (7)-(12) captures the additional belief biases resulting from receiving incorrect signals, controlling for potential impacts of TM1's type. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 8b.

Table A.7: Biases in beliefs about the average productivity in Experiment A

	DV: Abs. Errors in Log of Average Beliefs					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Lowest	0.352***	0.307***	0.382**	0.204***	0.348**	0.213***
	(0.133)	(0.074)	(0.156)	(0.077)	(0.140)	(0.063)
Second-Lowest	0.190**	0.056	0.059	-0.010	0.135*	0.028
	(0.086)	(0.062)	(0.094)	(0.059)	(0.081)	(0.057)
Second-Highest	-0.024	-0.046	-0.069	-0.056	-0.043	-0.038
	(0.061)	(0.044)	(0.057)	(0.043)	(0.056)	(0.039)
Highest	-0.091*	-0.062	-0.097*	-0.057	-0.086*	-0.042
	(0.052)	(0.039)	(0.054)	(0.040)	(0.051)	(0.037)
Round	-0.000	-0.005*	-0.003	0.000	-0.001	-0.003***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.000)
Constant	0.351***	0.393***	0.475***	0.276***	0.379***	0.381***
	(0.044)	(0.034)	(0.104)	(0.089)	(0.039)	(0.031)
H_0 : $\text{Lowest}_{ITA} = \text{Lowest}_{GTA}$						
F-statistic		0.089		1.052		0.787
p-value		0.765		0.306		0.376
H_0 : $\text{Second-Lowest}_{ITA} = \text{Second-Lowest}_{GTA}$						
F-statistic		1.600		0.390		1.159
p-value		0.207		0.533		0.282
H_0 : $\text{Second-Highest}_{ITA} = \text{Second-Highest}_{GTA}$						
F-statistic		0.085		0.029		0.006
p-value		0.771		0.865		0.939
H_0 : $\text{Highest}_{ITA} = \text{Highest}_{GTA}$						
F-statistic		0.201		0.346		0.485
p-value		0.654		0.557		0.486
R^2	0.163	0.113	0.159	0.077	0.132	0.075
Total Observations	1019	2178	1000	2178	5071	10873
Num. of Individuals	102	218	102	218	102	218

Note: The dependent variable is the absolute distance between *Log of Average Beliefs*, defined by the log of beliefs about the average productivity of TM1 and TM2, and the true average productivity. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional belief biases with respect to the average productivity of TM1 and TM2 that *Lowest* types make compared to *Average* types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 9a.

Table A.8: Biases in beliefs about the average productivity in Experiment B

	DV: Abs. Errors in Log of Average Beliefs											
	Round 1-10		Round 41-50		Round 1-50		Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA	(7) ITA	(8) GTA	(9) ITA	(10) GTA	(11) ITA	(12) GTA
Incorrect Signal	0.087	0.067	0.137	0.069	0.115	0.073	0.125	0.093	0.161*	0.107*	0.144*	0.104*
	(0.078)	(0.065)	(0.094)	(0.068)	(0.077)	(0.060)	(0.079)	(0.056)	(0.086)	(0.063)	(0.074)	(0.053)
Round	0.002	-0.009***	0.003	-0.004	-0.001*	-0.003***	0.002	-0.009***	0.003	-0.004	-0.001*	-0.002***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.001)	(0.000)	(0.003)	(0.003)	(0.002)	(0.003)	(0.001)	(0.000)
TM1 Lowest							0.260*	0.227*	0.206	0.345***	0.211	0.292**
							(0.138)	(0.129)	(0.173)	(0.122)	(0.139)	(0.120)
TM1 Second-Lowest							0.117	0.158**	-0.140	0.191***	-0.019	0.175***
							(0.078)	(0.062)	(0.114)	(0.057)	(0.081)	(0.045)
TM1 Second-Highest							-0.009	-0.079	-0.181	0.025	-0.097	-0.011
							(0.061)	(0.062)	(0.111)	(0.047)	(0.066)	(0.040)
TM1 Highest							0.087	-0.005	-0.032	0.036	0.026	0.030
							(0.067)	(0.062)	(0.109)	(0.048)	(0.070)	(0.045)
Constant	0.352***	0.404***	0.153	0.424***	0.366***	0.357***	0.243***	0.341***	0.179	0.301**	0.333***	0.257***
	(0.040)	(0.038)	(0.112)	(0.148)	(0.037)	(0.031)	(0.062)	(0.051)	(0.150)	(0.151)	(0.051)	(0.036)
H_0 : Incorrect Signal_{ITA} = Incorrect Signal_{GTA}												
F-statistic		0.037		0.348		0.187		0.104		0.261		0.184
p-value		0.849		0.556		0.666		0.748		0.610		0.669
R^2	0.010	0.013	0.019	0.008	0.017	0.019	0.065	0.110	0.110	0.133	0.075	0.117
Total Observations	980	917	966	929	4854	4627	980	917	966	929	4854	4627
Num. of Individuals	99	92	97	93	99	93	99	92	97	93	99	93

Note: The dependent variable is the absolute distance between *Log of Average Beliefs*, defined by the log of beliefs about the average productivity of TM1 and TM2, and the true average productivity. *Incorrect Signal* is the indicator of subjects who receive incorrect signals about TM1. The coefficient indicate the causal effect of receiving incorrect signals about TM1 on belief biases. The coefficient of *Incorrect Signal* captures the additional belief biases resulting from receiving incorrect signals. The coefficient of *Incorrect Signal* in Column (1)-(6) captures the additional belief biases resulting from receiving incorrect signals. The coefficient of *Incorrect Signal* in Column (7)-(12) captures the additional belief biases resulting from receiving incorrect signals, controlling for potential impacts of TM1's type. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 9b.

Section 4.1.3: Allocation choices

Table A.9: Absolute errors in choices in Experiment A

	DV: Abs. Errors in Choices					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Lowest	0.606	0.090	1.143	4.470*	1.126	2.204
	(2.740)	(1.664)	(2.875)	(2.466)	(2.785)	(1.538)
Second-Lowest	3.983	1.015	0.910	-0.412	1.859	1.099
	(3.068)	(1.869)	(2.373)	(2.566)	(2.606)	(2.223)
Second-Highest	-0.244	-0.190	-1.297	-2.751*	-1.100	-1.134
	(2.601)	(1.956)	(2.293)	(1.619)	(2.208)	(1.440)
Highest	-0.854	-1.251	-2.279	-1.900	-1.970	-1.455
	(2.467)	(1.421)	(1.629)	(1.570)	(1.906)	(1.260)
Round	-0.097	-0.096	-0.070	-0.048	-0.031	-0.004
	(0.157)	(0.102)	(0.118)	(0.065)	(0.023)	(0.017)
Constant	6.507***	7.092***	8.676	8.829***	7.126***	6.230***
	(1.719)	(1.218)	(5.490)	(3.038)	(1.606)	(1.133)
H_0 : Lowest$_{ITA}$ = Lowest$_{GTA}$						
F-statistic		0.026		0.776		0.115
p-value		0.872		0.379		0.734
H_0 : Second-Lowest$_{ITA}$ = Second-Lowest$_{GTA}$						
F-statistic		0.687		0.144		0.049
p-value		0.408		0.705		0.824
H_0 : Second-Highest$_{ITA}$ = Second-Highest$_{GTA}$						
F-statistic		0.000		0.270		0.000
p-value		0.987		0.604		0.989
H_0 : Highest$_{ITA}$ = Highest$_{GTA}$						
F-statistic		0.020		0.028		0.051
p-value		0.889		0.866		0.821
R^2	0.014	0.004	0.015	0.039	0.012	0.013
Total Observations	1019	2178	1000	2178	5071	10873
Num. of Individuals	102	218	102	218	102	218

Note: The dependent variable is *Abs. Choice Errors*, defined by the absolute difference between the subjectively optimal action and their actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional choice errors that *Lowest* types make compared to *Average* types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 12a.

Table A.10: Absolute errors in choices in Experiment B

	DV: Abs. Errors in Choices					
	Round 1-10		Round 41-50		Round 1-50	
	(1) ITA	(2) GTA	(3) ITA	(4) GTA	(5) ITA	(6) GTA
Lowest	8.379*	-1.079	4.849	-1.689	6.912**	-0.735
	(4.484)	(1.890)	(3.286)	(2.782)	(2.882)	(2.008)
Second-Lowest	-0.668	0.249	4.692	-4.257*	3.279	-1.738
	(2.841)	(1.656)	(3.671)	(2.233)	(2.926)	(1.683)
Second-Highest	-2.485	0.141	0.714	-1.234	1.409	-0.205
	(2.745)	(1.866)	(2.735)	(2.629)	(2.381)	(2.038)
Highest	-3.180	0.252	0.941	-3.610	0.365	-1.897
	(2.574)	(1.556)	(2.861)	(2.390)	(2.087)	(1.782)
Round	0.214	0.234*	0.027	-0.012	0.029	0.013
	(0.136)	(0.128)	(0.164)	(0.100)	(0.027)	(0.018)
Constant	5.772**	2.527**	4.120	7.110	4.289**	4.848***
	(2.464)	(1.228)	(7.645)	(4.556)	(1.719)	(1.427)
H_0 : Lowest$_{ITA}$ = Lowest$_{GTA}$						
F-statistic		3.797		2.319		4.765
p-value		0.053		0.129		0.030
H_0 : Second-Lowest$_{ITA}$ = Second-Lowest$_{GTA}$						
F-statistic		0.078		4.364		2.220
p-value		0.780		0.038		0.138
H_0 : Second-Highest$_{ITA}$ = Second-Highest$_{GTA}$						
F-statistic		0.629		0.265		0.267
p-value		0.429		0.607		0.606
H_0 : Highest$_{ITA}$ = Highest$_{GTA}$						
F-statistic		1.308		1.499		0.683
p-value		0.254		0.222		0.410
R^2	0.086	0.007	0.018	0.033	0.031	0.008
Total Observations	980	917	966	929	4854	4627
Num. of Individuals	99	92	97	93	99	93

Note: The dependent variable is *Abs. Choice Errors*, defined by the absolute difference between the subjectively optimal action and their actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across subject types in comparison to *Average* types (the omitted category). For example, the coefficient of *Lowest* captures the additional choice errors that *Lowest* types make compared to *Average* types. ITA stands for Individual Task Assignment. GTA stands for Group Task Assignment. Standard errors clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 12b.

Table A.11: Treatment effects in Experiment A

	DV: Output loss (%)					
	Individual Task Assignment			Group Task Assignment		
	(1) Round 1-10	(2) Round 41-50	(3) Round 1-50	(4) Round 1-10	(5) Round 41-50	(6) Round 1-50
Lowest	8.636***	4.805*	6.355***	5.615***	0.475	2.641**
	(3.027)	(2.760)	(2.188)	(1.800)	(1.015)	(1.056)
Second-Lowest	5.424**	3.003	3.271*	-0.460	-0.335	0.236
	(2.429)	(2.136)	(1.895)	(1.244)	(1.145)	(0.954)
Second-Highest	-1.164	-2.324*	-1.470	-1.026	-1.899***	-0.829
	(2.832)	(1.214)	(1.515)	(1.365)	(0.684)	(0.704)
Highest	-1.707	-0.627	-1.134	-1.953*	-1.517**	-0.755
	(2.247)	(1.456)	(1.573)	(1.094)	(0.687)	(0.685)
Round	-0.586***	0.080	-0.167***	-0.296***	-0.061	-0.105***
	(0.179)	(0.140)	(0.020)	(0.099)	(0.051)	(0.011)
Constant	13.334***	1.054	11.686***	8.924***	6.765***	7.623***
	(2.375)	(6.384)	(1.372)	(1.061)	(2.458)	(0.601)
R^2	0.065	0.041	0.059	0.045	0.013	0.031
Total Observations	1020	1000	5080	2180	2180	10900
Num. of Individuals	102	102	102	218	218	218

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences among subject types compared to *Average* types (the omitted category). For instance, the coefficient of *Lowest* represents the additional percentage points of output loss incurred by *Lowest* types in comparison to Average types. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 14.

Table A.12: Allocative efficiency across subject types in Experiment B

	DV: Output loss (%)					
	Individual Task Assignment			Group Task Assignment		
	(1) Round 1-10	(2) Round 41-50	(3) Round 1-50	(4) Round 1-10	(5) Round 41-50	(6) Round 1-50
Lowest	-3.478	0.022	-1.831	-0.901	0.870	0.352
	(4.210)	(3.640)	(3.365)	(1.948)	(1.646)	(1.296)
Second-Lowest	-3.302	1.824	-1.214	2.076	2.855	2.287
	(4.444)	(3.477)	(3.497)	(2.007)	(1.749)	(1.459)
Second-Highest	-0.226	-0.469	-0.535	0.809	-0.890	0.918
	(4.523)	(3.616)	(3.440)	(1.984)	(1.076)	(1.278)
Highest	0.148	1.399	0.525	0.795	-0.935	0.339
	(4.399)	(3.258)	(3.353)	(2.009)	(1.108)	(1.124)
Round	-0.280	0.176	-0.166***	-0.150	-0.165*	-0.062***
	(0.201)	(0.154)	(0.025)	(0.145)	(0.091)	(0.015)
Constant	16.956***	-1.759	14.190***	6.551***	11.024**	5.754***
	(4.116)	(7.544)	(3.197)	(1.951)	(4.624)	(1.093)
R^2	0.012	0.006	0.029	0.008	0.030	0.013
Total Observations	990	990	4950	930	930	4650
Num. of Individuals	99	99	99	93	93	93

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences among subject types compared to *Average* types (the omitted category). For instance, the coefficient of *Lowest* represents the additional percentage points of output loss incurred by *Lowest* types in comparison to Average types. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 15.

Table A.13: Allocative efficiency across TM1's types in Experiment B

	DV: Output loss (%)					
	Individual Task Assignment			Group Task Assignment		
	(1) Round 1-10	(2) Round 41-50	(3) Round 1-50	(4) Round 1-10	(5) Round 41-50	(6) Round 1-50
TM1 Lowest	2.964	5.013	4.321	3.287	4.426**	3.501**
	(4.236)	(3.506)	(2.789)	(2.098)	(1.899)	(1.697)
TM1 Second-Lowest	-3.328	-2.700	-2.530	3.091**	2.699*	2.448**
	(3.567)	(2.610)	(2.014)	(1.401)	(1.423)	(1.190)
TM1 Second-Highest	3.356	1.655	2.241	0.724	0.533	0.268
	(4.000)	(3.043)	(2.471)	(1.564)	(1.366)	(1.335)
TM1 Highest	-1.489	0.431	-0.743	1.689	0.369	0.888
	(3.260)	(2.806)	(1.932)	(1.391)	(1.312)	(1.241)
Round	-0.280	0.176	-0.166***	-0.150	-0.165*	-0.062***
	(0.201)	(0.154)	(0.025)	(0.145)	(0.091)	(0.015)
Constant	15.485***	-1.855	13.115***	5.446***	9.901**	5.208***
	(3.102)	(6.701)	(1.745)	(1.266)	(4.196)	(1.004)
R^2	0.025	0.036	0.052	0.012	0.032	0.020
Total Observations	990	990	4950	930	930	4650
Num. of Individuals	99	99	99	93	93	93

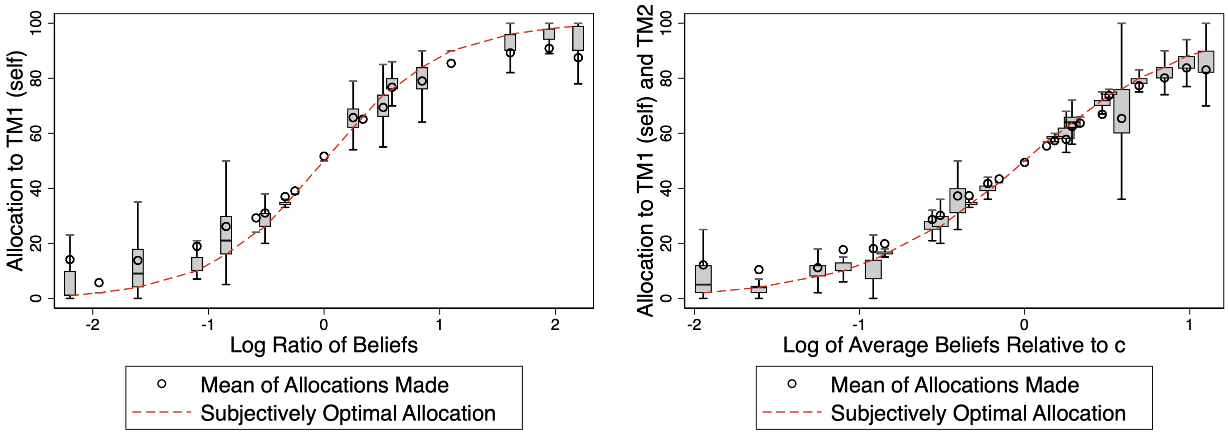
Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *TM1 Lowest*, *TM1 Second-Lowest*, *TM1 Second-Highest*, and *TM1 Highest* refer to the indicators of TM1's types with whom the subject is matched. The coefficients indicate the differences among subjects matched with each type of TM1 compared to those whose TM1 is a *Average* type (the omitted category). For instance, the coefficient of *TM1 Lowest* represents the additional percentage points of output loss incurred by subjects whose TM1 is *Lowest* types in comparison to those whose TM1 is *Average* types. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 16.

Table A.14: Treatment effects in Experiment B

	DV: Output loss (%)					
	Individual Task Assignment			Group Task Assignment		
	(1) Round 1-10	(2) Round 41-50	(3) Round 1-50	(4) Round 1-10	(5) Round 41-50	(6) Round 1-50
Incorrect Signal	9.383***	5.457**	6.082***	0.074	1.577	1.248
	(2.831)	(2.359)	(2.026)	(1.227)	(1.509)	(1.174)
Round	-0.280	0.176	-0.166***	-0.150	-0.165*	-0.062***
	(0.201)	(0.154)	(0.025)	(0.145)	(0.091)	(0.015)
Constant	12.993***	-2.546	11.951***	7.259***	11.034**	6.316***
	(1.432)	(6.708)	(0.963)	(1.041)	(4.209)	(0.584)
R^2	0.061	0.033	0.059	0.001	0.008	0.010
Total Observations	990	990	4950	930	930	4650
Num. of Individuals	99	99	99	93	93	93

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). *Incorrect Signal* is the indicators of of receiving incorrect signals about TM1. The coefficients indicate the causal effect of receiving incorrect signal about TM1 on output loss. We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance. See also Figure 17.

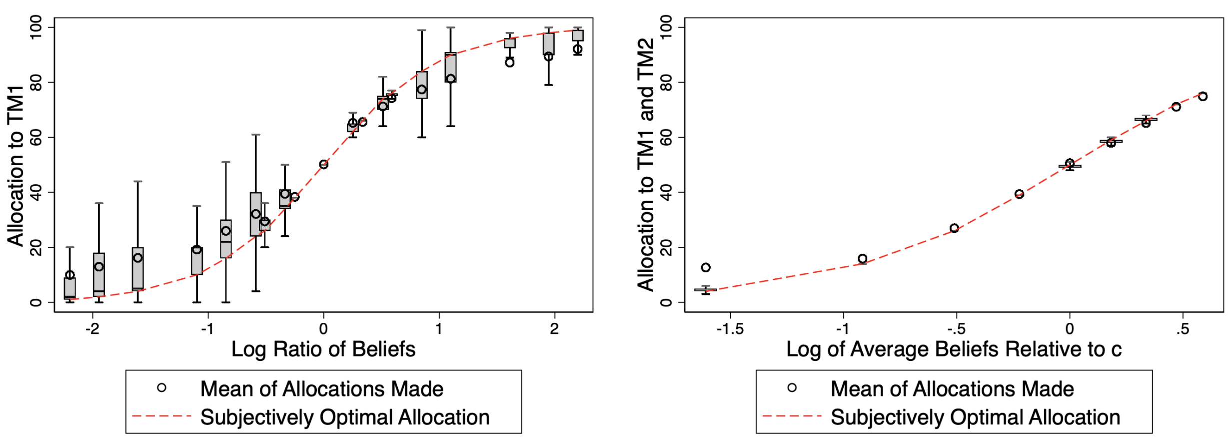
B Additional figures and tables



(a) Individual Task Assignment (b) Group Task Assignment

Figure B.1: Subjectively optimal choices in Experiment A

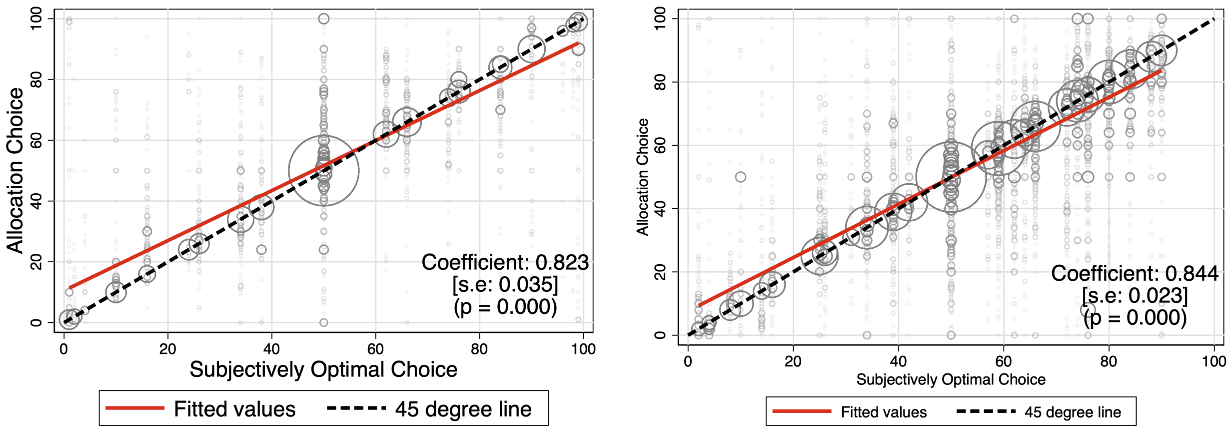
Note: Panel (a) displays box whisker plots of allocation choices conditional on beliefs about the productivity ratio between TM1 (self) and TM2. Panel (b) displays box whisker plots of allocation choices conditional on beliefs about the average productivity of TM1 and TM2 relative to the productivity of a robot $c \in \{30, 50, 70\}$. In both, the means of allocation choices made are overlaid on the box whisker plots and marked as a circle. The red dashed line represents the theoretical benchmark. If one optimizes an allocation given their, potentially biased, beliefs, the allocation must be on the red dashed line.



(a) Egoless Individual Task Assignment (b) Egoless Group Task Assignment

Figure B.2: Subjectively optimal choices in Experiment B

Note: Panel (a) displays box whisker plots of allocation choices conditional on beliefs about the productivity ratio between TM1 and TM2. Panel (b) displays box whisker plots of allocation choices conditional on beliefs about the average productivity of TM1 and TM2 relative to the productivity of a robot $c \in \{50\}$. In both, the means of allocation choices made are overlaid on the box whisker plots and marked as a circle. The red dashed line represents the theoretical benchmark. If one optimizes an allocation given their, potentially biased, beliefs, the allocation must be on the red dashed line.

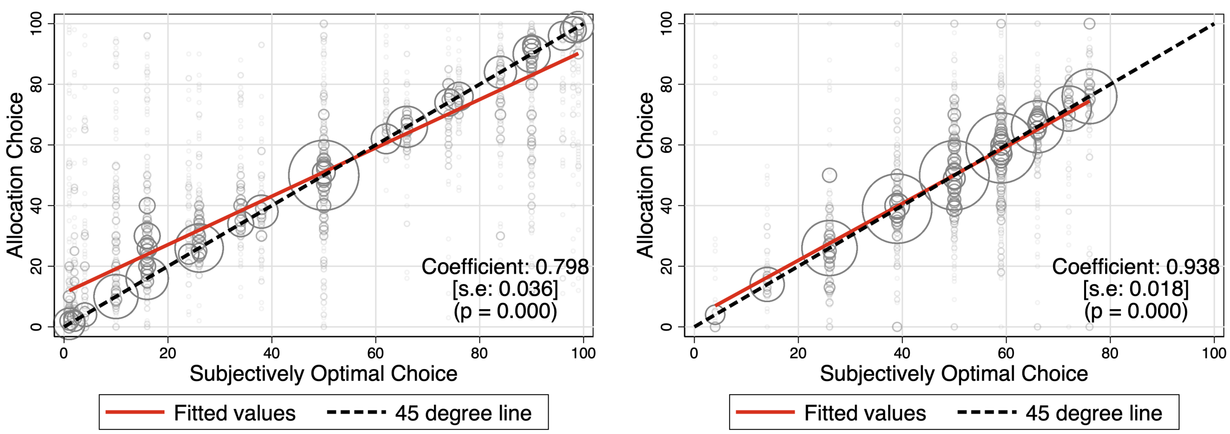


(a) Individual Task Assignment

(b) Group Task Assignment

Figure B.3: Linear fit of allocation choice on subjectively optimal choice in Experiment A

Note: We regress choice allocations on the theoretical benchmark that subjects should choose in order to maximize their perceived expected output based on their beliefs. The figure shows the regression results for Experiment 1 overlaid with the scatter plots of allocation choices relative to the theoretical benchmark. The size of the markers indicates the frequency of observations. The red solid line represents the fitted values of the regression estimates, while the black dashed line represents the 45-degree line as the benchmark.



(a) Individual Task Assignment

(b) Group Task Assignment

Figure B.4: Linear fit of allocation choice on subjectively optimal choice in Experiment B

Note: We regress choice allocations on the theoretical benchmark that subjects should choose in order to maximize their perceived expected output based on their beliefs. The figure shows the regression results for Experiment 2 overlaid with the scatter plots of allocation choices relative to the theoretical benchmark. The size of the markers indicates the frequency of observations. The red solid line represents the fitted values of the regression estimates, while the black dashed line represents the 45-degree line as the benchmark.

Table B.1: Treatment effects in Experiment A (fully saturated)

	DV: Output Loss (%)		
	Round 1-10	Round 41-50	Round 1-50
	(1)	(2)	(3)
Lowest	8.636***	4.806*	6.354***
	(3.014)	(2.749)	(2.180)
Second-Lowest	5.424**	3.004	3.271*
	(2.419)	(2.127)	(1.889)
Second-Highest	-1.164	-2.336*	-1.464
	(2.820)	(1.208)	(1.510)
Highest	-1.707	-0.625	-1.135
	(2.238)	(1.450)	(1.567)
GTA	-2.815	-0.678	-2.491*
	(2.101)	(1.180)	(1.283)
Lowest $\times GTA$	-3.021	-4.332	-3.713
	(3.510)	(2.931)	(2.422)
Second-Lowest $\times GTA$	-5.884**	-3.339	-3.035
	(2.720)	(2.416)	(2.116)
Second-Highest $\times GTA$	0.139	0.437	0.635
	(3.133)	(1.388)	(1.665)
Highest $\times GTA$	-0.246	-0.892	0.380
	(2.491)	(1.605)	(1.710)
Round	-0.388***	-0.017	-0.125***
	(0.088)	(0.056)	(0.010)
Constant	12.247***	5.474**	10.613***
	(2.045)	(2.693)	(1.232)
Total Observations	3200	3180	15980
Num. of Individuals	320	320	320

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). We control for rounds, and standard errors are clustered at the individual level and are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance.

Table B.2: Treatment effects in Experiment B (fully saturated)

	DV: Output Loss (%)		
	Round 1-10	Round 41-50	Round 1-50
	(1)	(2)	(3)
Incorrect signal	9.383***	5.457**	6.082***
	(2.824)	(2.353)	(2.021)
GTA	-5.015***	-1.924*	-2.987***
	(1.160)	(1.014)	(0.830)
Incorrect signal×GTA	-9.309***	-3.880	-4.834**
	(3.078)	(2.793)	(2.336)
Round	-0.217*	0.011	-0.116***
	(0.125)	(0.091)	(0.015)
Constant	12.645***	4.963	10.668***
	(1.169)	(3.958)	(0.817)
Total Observations	1920	1920	9600
Num. of Individuals	192	192	192

Note: The dependent variable is *Output loss (%)*, defined by the additional potential output that a subject could have achieved if they had made optimal decisions given the (true) productivity of TM1 and TM2. Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance.