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

Marco Castillo[†] Sora Youn[‡]

November 7, 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 are grateful to Alex Brown, Catherine Eckel, PJ Healy, Paolina Median, Ragan Petrie, Danila Serra, Rodrigo Velez, and 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. We thank Department of Economics, Texas A&M Univerisity for financial support. 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 are widespread. A vast body of literature has documented belief biases in various domains and examined their economic consequences.[1] A central theme in the literature is why and how biases persist.[2] This paper concentrates instead on 'when' biased beliefs are harmful or innocuous. In particular, we experimentally test a theoretical condition in which biased beliefs can lead to optimal action in the context of task allocations within teams (Heidhues, Kőszegi, and Strack, 2018).

There are at least three reasons why the answer to this question is important. Firstly, past work suggests that an accurate understanding of each team member's ability and strengths is important for assigning tasks strategically and making the team perform better (Bandiera, Barankay, and Rasul, 2007; 2009, Burgess, Propper, Ratto, Kessler Scholder, and Tominey, 2010, Delfgaauw, Dur, and Souverijn, 2020). However, in many cases, direct observation and assessment of individual productivity are lacking. In these situations, task allocation

[1]For example, studies have documented persistent biases in self-perception, which may be either overly positive or overly negative. Some individuals often maintain an excessively optimistic view of themselves (for a review, Santos-Pinto and de la Rosa, 2020). Others 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).

[2]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). Conversely, 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 (e.g., Kendall and Oprea, forthcoming).

depends on the subjective, potentially biased beliefs of the task allocator about each team member's productivity. Moreover, learning of each team member's productivity may be impaired by the fact that only imperfectly measured aggregate team output is observed.

Second, biased beliefs can be persistent and challenging to rectify. People may ignore or misinterpret information (Epley and Gilovich, 2016; Golman, Hagmann, and Loewenstein, 2017; Rabin and Schrag, 1999) and selectively recall memories (Huffman, Raymond, and Shvets, 2022; Zimmermann, 2020) to maintain their existing and preferred beliefs. It is also possible that people fail to recognize their mistakes despite rich data (Enke, 2020; Enke and Zimmermann, 2019; Graeber, 2022; Hanna, Mullainathan, and Schwartzstein, 2014). In the context of teamwork, these tendencies can lead the task allocator to (mis)attribute the observed team output in a way that is consistent with her (biased) prior beliefs (Coutts, Gerhards, and Murad, 2020). The task allocator may not notice that the team could be better off if she allocated tasks differently among team members (Dargnies, Hakimov, and Kübler, 2022). In such cases, attempts to correct belief biases may be ineffective (Haaland, Roth, and Wohlfart, 2023; Thaler, Sunstein, and Balz, 2014).

Third, understanding the conditions that promote economically desirable outcomes can help in devising strategies to counteract belief biases without correcting biases directly. The knowledge allows us to design institutions, such as incentives, information structures, and more comprehensive organizational frameworks, that can protect against the economic harm caused by biased beliefs (see Enke, Graeber, and Oprea, 2023; Recalde and Vesterlund, 2022, for recent examples).

We build on the work of Heidhues et al. (2018) who characterize conditions for efficient task allocation in the presence of biased beliefs about team members' productivity. To illustrate the theoretical framework in an environment similar to our experiment, consider a manager and two team members: Teammate 1 (TM1) and Teammate 2 (TM2). The manager observes the team's output which depends on each member's productivity, the manager's task allocation choice, and a random component. The objective of the manager is to maximize the expected output by allocating tasks to TM1 and TM2. In the long run, the manager's beliefs converge to stable beliefs that satisfy two conditions: i) there exists an optimal task allocation for maximizing expected output conditional on stable beliefs, referred to as perceived expected output; ii) this allocation, on average, achieves actual output that matches the perceived expected output.[3] Crucially, stable beliefs can be biased. If the observed output is unsurprising to the manager on average, the manager finds no reason to revise her beliefs and the corresponding allocation, leading her to falsely conclude that her biased beliefs about each team member's productivity are correct.

Stable yet biased beliefs, therefore, can be harmful if the task allocation that maximizes the perceived expected output is suboptimal given the true productivities of TM1 and TM2. Suppose the optimal task allocation requires knowledge of the productivity ratio between TM1 and TM2.[4] In this environment, which we refer to as Individual Task Assignment (ITA), biased beliefs about each team member's productivity results in suboptimal allocation in the long run unless the manager has correct beliefs about the productivity ratio. Conversely, in cases where optimal allocation only requires knowing the average productivity of TM1 and TM2, referred to as Group Task Assignment (GTA), biased beliefs about each team member's productivity can be inconsequential.[5] The manager may misattribute, but such misattributions do not distort the optimal allocation as long as the manager holds correct beliefs about the average productivity of TM1 and TM2. In summary, Heidhues et al. (2018) predicts that the optimal allocation of tasks is more likely in Group Task Assignment (GTA) than in Individual Task Assignment (ITA) in the presence of similarly biased beliefs.[6]

[3] The equilibrium concept is known as the Berk-Nash equilibrium. For more details, see Esponda and Pouzo (2016) and Fudenberg, Lanzani, and Strack (2021).

[4] An example of this scenario is when the manager allocates tasks between TM1 and TM2. By assigning more tasks to the more productive team member, team output can be maximized.

[5] An example of this scenario is when the manager allocates tasks to both TM1 and TM2 as a group, while also considering another team with known productivity, such as an outsourcing team. The manager needs to learn how productive the two members are on average compared to the outsourcing team. Notice that beliefs about the average productivity can be correct even when the manager holds incorrect beliefs about each team member. For example, if the manager overestimates TM1's productivity and underestimates TM2's, these biases can offset each other without affecting the learning of the average productivity.

[6] Heidhues et al. (2018) call Individual Task Assignment *Identifiability* and Group Task Assignment *Non-*

Testing this hypothesis in the laboratory poses several experimental design challenges. First, and most importantly, it requires creating an environment in which persistent belief biases are present, whether naturally occurring or experimentally induced. Naturally occurring biases offer real-life relevance, while experimentally induced biases allow for the establishment of causal relationships. Second, the design should enable the measurement of beliefs. Observing beliefs allows us to assess whether subjects respond to incentives. Given the way Individual Task Assignment (ITA) and Group Task Assignment (GTA) are structured, we expect that subjects in ITA would be more engaged in learning the productivity ratio between TM1 and TM2, while those in the GTA group would be more engaged in learning the average productivity of TM1 and TM2. Furthermore, it allows us to estimate subjects' perceptions so that we can determine whether beliefs converge to stable beliefs. It also helps us decompose treatment effects driven by biased beliefs from choice errors – those choices that are inconsistent with their beliefs.

To achieve these goals, we design an experiment where every subject faces an individual decision-making problem: task allocation with two fixed team members whose productivities are unknown.[7] Our experimental design varies along two dimensions: whether optimal allocation requires knowledge of the productivity ratio between the team members (Individual Task Assignment, ITA) or the average productivity of the team members (Group Task Assignment, GTA), and whether biased beliefs are naturally occurring (Experiment A) or experimentally induced (Experiment B). See Figure 2 for the overview of the experimental design.

In Experiment A (naturally occurring beliefs), each subject is assigned the role of TM1 and is paired with another subject (TM2). Based on extensive research on the better-than-average phenomenon (e.g., for a review, see Zell, Strickhouser, Sedikides, and Alicke,

identifiability.

[7]To eliminate strategic concerns and interactions between team members, the productivity of each team member is pre-determined by a real-effort task (Gill and Prowse, 2012; Gill and Prowse, 2019; Chen and Schildberg-Hörisch, 2019).

2020), we conjecture that subjects with below-average productivity likely hold incorrect and persistent beliefs about TM1 (themselves). In Experiment B (experimentally induced beliefs), we introduce an exogenous shock in beliefs. Each subject is paired with two others, one labeled as TM1 and the other as TM2. Subjects receive a random signal indicating whether or not TM1's productivity is below average with a 75% accuracy. Regardless of subjects' prior beliefs about TM1, this procedure causally induces greater belief biases for those receiving incorrect signals.

Both Experiment A and Experiment B have two treatments: Individual Task Assignment (ITA) and Group Task Assignment (GTA). In both treatments, subjects state how to divide 100 hypothetical projects between two parties. In ITA, this allocation is between TM1 and TM2. In GTA, it is between the group consisting of TM1 and TM2 and a robot player with known productivity. Note that the treatments differ only in the determinant for the optimal allocation. In ITA, the crucial factor is understanding the productivity ratio between TM1 and TM2. In GTA, understanding the average productivity of TM1 and TM2 is sufficient. Subjects repeat the allocation task for 50 rounds. In every round, subjects encounter an experimental interface where they separately report their beliefs about each team member's productivity and their allocation choices. At the end of every round, subjects observe the team's output, further confounded by the addition of random noise. They do not observe the individual contributions of each team member in any treatment.

We find evidence supporting the theoretical prediction (presented in Section 5.2). Despite similarly biased beliefs, greater allocative efficiency is achieved in Group Task Assignment (GTA) compared to Individual Task Assignment (ITA). This is true whether beliefs are home-grown or experimentally induced. In Experiment A, the average output loss is almost 65 percent larger in ITA than GTA in the last ten rounds (5.3 v. 3.3, $p < 0.01$). In the last ten rounds of Experiment B, the average output loss is about 75 percent larger in ITA than GTA (6.9 v. 3.9, $p < 0.01$).

We also show that the treatment effects are mainly driven by biased beliefs. In Section

5.1.1, we confirm that subjects respond to incentives. Subjects in ITA have better-calibrated beliefs about the productivity ratio between TM1 and TM2 and subjects in GTA have better-calibrated beliefs about the average productivity of TM1 and TM2. Section 5.1.2 verifies our experimental manipulation. Subjects holding the most biased beliefs are those with the lowest productivity in Experiment A and those receiving incorrect signals in Experiment B. Importantly, we find biased beliefs hinder learning about the productivity ratio between TM1 and TM2 more than they hurt the learning of the average productivity of TM1 and TM2. Therefore, such biased beliefs have less harmful effects on allocative efficiency in GTA. The decomposition of the treatment effect confirms that biased beliefs about the productivity ratio and biased beliefs about the average productivity are the primary driving factors in ITA and GTA, respectively (see Section 5.2.4). While subjects may not optimize task allocation choices contingent on their beliefs, our experimental results indicate that the allocation choices in our experiments largely align with subjects' beliefs, resulting in minimal impact on efficiency (Section 5.1.3).[8]

Next, we examine whether the biased beliefs at play resemble *noisy* equilibrium play.[9] Recall that beliefs converge into stable states in the long run, where the expected output, conditioned on these beliefs, aligns with the actual expected output. On average, we observe a 10-20% decrease in the expectation-reality gap between the first and last ten rounds in both treatments ($p < 0.01$). The remaining gap may indicate that beliefs are off-equilibrium or that subjects tolerate expectation-reality gaps.

To assess this point, we explore whether there are differences in expectation-reality gaps, either based on subject types in Experiment A or the correctness of signals in Experiment B. If it is off-equilibrium, we anticipate those who were further from equilibrium in the beginning would likely remain so until the end. We find mixed evidence consistent with a

[8]This is partly attributed to our interface's ability to improve decision quality based on subjects' beliefs.

[9]Marray, Krishna, and Tang (2020) and Goette and Kozakiewicz (2022) provide mixed evidence on whether subjects play Berk-Nash equilibrium. Our focus is different from theirs. Our main research question is the interaction between institutional characteristics and optimal behavior in the presence of biased beliefs, especially in the context of task allocations within teamwork.

tolerance of expectation-reality gaps. In Experiment A, we observe significant heterogeneity by subject types in the first 10 rounds (p=0.058 for ITA and p=0.047 for GTA) but not in the last ten rounds (p >0.1 for both ITA and GTA). In Experiment B, subjects receiving incorrect signals consistently experience a greater gap during the experiment in ITA, but the gaps are not significantly different in GTA.

Our paper contributes to the literature on organizational economics. While substantial theoretical research has focused on identifying decision rules for optimal delegation (for a review, Bendor, Glazer, and Hammond, 2001), there is still a scarcity of empirical evidence regarding their impact on organizational decision quality (e.g., Dlugosz, Gam, Gopalan, and Skrastins, Forthcoming). This paper emphasizes the significance of organizational design in task allocation and its influence on the behavior and performance of organizations.

Our findings are in line with a few papers that explore institutional design to mitigate the adverse impacts of incorrect or misspecified mental models (Enke et al., 2023; Recalde and Vesterlund, 2022). Esponda, Vespa, and Yuksel (Forthcoming) and Gagnon-Bartsch, Rabin, and Schwartzstein (2021) suggest that providing limited information can promote optimal behavior by encouraging people to pay closer attention to data. However, our results show that redesigning incentives can similarly lead to optimal behavior, even with the same amount of information available. Our results extend beyond teamwork. Discrimination in the workplace is an obvious application (e.g., Neumark, 2018). Furthermore, there are numerous situations where conflicts can arise due to incorrect beliefs and subsequent misattribution (e.g., Bertrand and Mullainathan, 2001; Wolfers, 2002). A remaining question is whether these institutional changes are feasible.

The paper is organized as follows. Section 2, we introduce the theoretical framework. Section 3 outlines the experimental design. In Section 4, we present the hypotheses. Section 5 provides the results. Finally, in Section 6, we offer our concluding remarks.

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) show that the DM's beliefs b_1 and b_2 are stable when the DM maximizes the (perceived) expected output given b_1 and b_2, and if the corresponding (subjectively) optimal action x^e produces the actual expected output $\mu(x^e|a_1, a_2)$ that is identical to the perceived expected output $\mu(x^e|b_1, b_2)$.[10] If the DM holds stable beliefs, she finds no inconsistency in her beliefs because her expectations, given her possibly biased beliefs, match reality. Therefore, the DM's beliefs converge to stable beliefs in the long run, if such beliefs exist.

Definition 1. The DM's beliefs b_1 and b_2 are stable if

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

[10] Heidhues et al. (2018) define *Surprise* as the difference between the actual expected output and her perceived expected output. The actual expected output is not surprising if beliefs are stable: *Surprise* $= \mu(x|a_1, a_2) - \mu(x|b_1, b_2) = 0$. The stable beliefs and the corresponding optimal action constitute a Berk-Nash equilibrium.

Stable beliefs can lead to suboptimal actions that do not maximize the actual expected output if they are biased. Heidhues et al. (2018) demonstrate that these biased stable beliefs are not detrimental and lead to the maximization of the actual expected output if μ depends solely on a summary statistic of a_1 and a_2, rather than 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.

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

- **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.

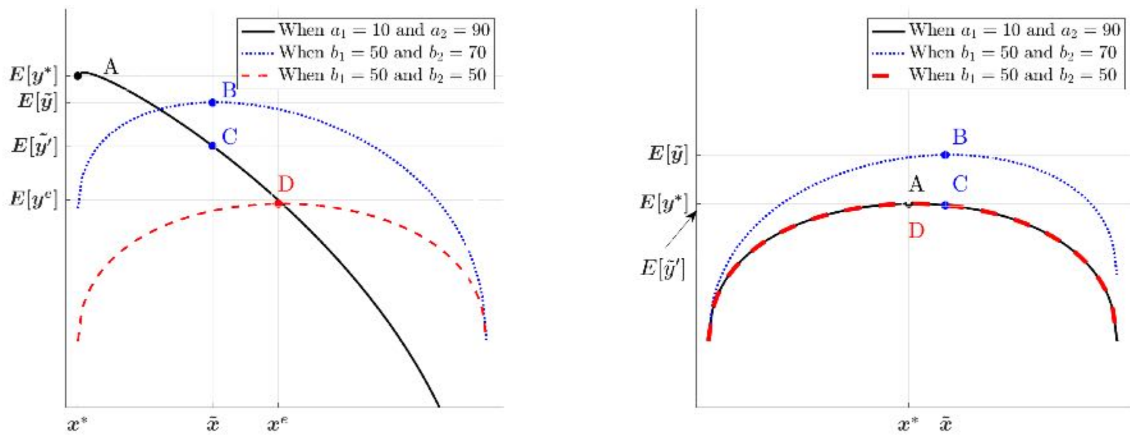
$$\mu^{GTA}(x|a_1, a_2) = \frac{a_1 + a_2}{2}\sqrt{x} + 50^{[11]}\sqrt{100 - x} \tag{2}$$

2.2 Illustrative example

To elaborate stable yet 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. Note that we assume the DM assigns zero probability to the true value of a_1. In other words, her beliefs are *misspecified*. Heidhues et al. (2018) further assumes that the DM's beliefs about a_1 are degenerated, with her being 100% certain about her incorrect belief regarding a_1. We relax this assumption by allowing beliefs about a_1 to vary (e.g., Esponda and Pouzo, 2016; Fudenberg et al., 2021).

Panel (a) of Figure 1 shows the case of Individual Task Assignment (ITA). The black solid

[11]To test for the robustness of the results, we vary this parameter to 30 and 70 in the experiment.



(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)).

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 signals to the DM that her beliefs are wrong, 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). 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 that is not surprising to the DM. 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 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, to her surprise, the DM observes the actual expected output $E[\tilde{y}'] = \mu^{GTA}(\tilde{x}|a_1 = 10, a_2 = 90)$ (Point C). From the gap between the realization and expectation, the DM realizes that her beliefs are wrong, specifically that her expectation was too high.

In response, she adjusts her beliefs. 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, under this wrong beliefs, her perceived expected output aligns exactly with the actual expected output. This is because the DM holds correct beliefs about the average of a_1 and a_2 even though she has incorrect beliefs about each individually. Therefore, Point D represents the stable yet biased beliefs that achieve the highest possible output.

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 $a_1 = a_2$.[12] 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 and 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 while holding biased beliefs about each of the parameters.

For simplicity, we have illustrated beliefs b_1 and b_2 as point beliefs. Proposition 1 relaxes this assumption. Suppose the DM forms belief distributions about a_1 and a_2. Denote by b_1 and b_2 the expected value of each belief distribution. 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,

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

called the stable action, is always optimal; the stable action maximizes the actual expected output in Group Task Assignment (GTA). However, in Individual Task Assignment (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 A □

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 bias Zell et al. (2020), we conjecture that subjects with below-average productivity will likely hold incorrect and persistent beliefs about TM1 (themselves). Second, we introduce random variations in belief biases. In Experiment B, 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). Screenshots of the decision-making screens can be found in Section H.

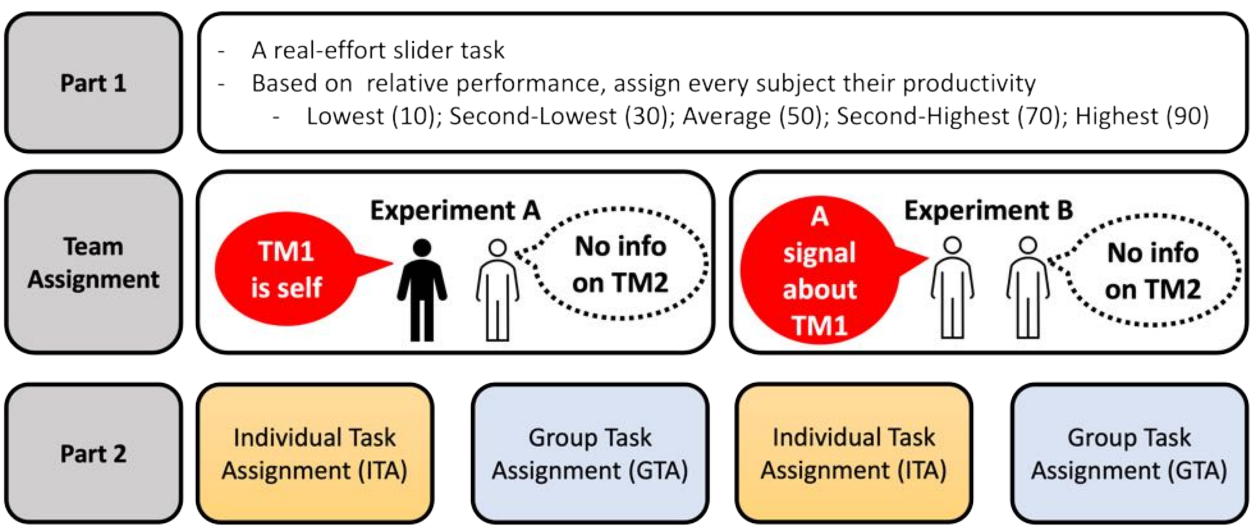


Figure 2: Overview of Experimental Design

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.

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 relative performance, their absolute performance in Part 1 is compared to a fixed group who completed the same slider task before the conduct of this study.[13] 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.

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

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.

[13] 222 subjects completed the same slider task in Fall 2021.

3.2 Team assignment

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

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 both in Experiment A and Experiment B.

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.[14]

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). Subjects are instructed to allocate 100 hypothetical projects between TM1

[14]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 B 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).

and TM2. In Group Task Assignment (GTA), the average of each team member's productivity determines the expected output, as shown in Equation (2). Subjects are told to allocate 100 hypothetical projects between a group of TM1 and TM2, and the robot player. In Experiment A, we vary the robot's productivity to 30, 50, or 70 to explore result dependency.[15] In Experiment B, the robot's productivity is set to be 50. In both ITA and GTA, subjects observe the realized outputs that are confounded by an additive 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.[16]

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. Yet, subjects are able to infer it from the simulator. 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

[15] In Experiment A, the expected output function for Group Task Assignment (GTA) is as follows: c is a constant, either 30, 50, or 70.

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

When we present the results, we pool the three versions of GTA, as the results are robust. Table H.13 shows that there is no significant difference in our efficiency measure across the three versions of GTA. In Column (2), we test whether the heterogeneity across subject types is significantly different across the three versions. We fail to reject a joint test of the interactions between the indicators for robot's productivity and the indicators for subject types being zero. For Robot type 30, the F-statistic is 1.89 (p=0.114). For Robot type 70, the F-statistic is 1.81 (p=0.128).

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

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 (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 in 2022. Subjects were recruited using ORSEE (Greiner, 2015). Table 1 summarizes the treatment conditions and the sample sizes. The experiment lasted, on average, one hour. Payments averaged approximately $15.24. Once subjects are seated in a computer station, an experimenter reads 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 reads them aloud. Subjects enter Part 2 simultaneously, 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.

Table 1: Treatment conditions and sample sizes

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

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

4 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 subjects would tend to believe their performance of the trivial real-effort task in Part 1 is at least as high as the average. Therefore, *Lowest* and *Second-Lowest* types, whose performance is below the average, would overestimate their relative performance and 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 are more likely to develop and maintain incorrect beliefs about TM1, regardless of their prior beliefs about TM1. 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. We define the absolute distance between the perceived expected output and the actual expected output, i.e., $|\mu(x|a_1, a_2) - \mu(x|b_1, b_2)|$, as *Abs. Surprise*. As beliefs converge to stable states, *Abs. Surprise* should approach zero. However, even if it does not, it may not necessarily indicate being off-equilibrium because people may have a degree of tolerance for positive expectation-reality gaps. To distinguish between off-equilibrium situations and subjects' tolerance, we can examine heterogeneity. If it is the latter, there should be no variation in *Abs. Surprise* across subject types in Experiment A or based on whether participants receive incorrect signals in Experiment B. An underlying assumption is that the tolerance level is not correlated with the degree of belief biases.

Hypothesis 3.1. (Stable beliefs) *Abs. Surprise* decreases over time and approach zero.

Hypothesis 3.2. (Heterogeneity in expectation-reality gap) In Experiment A, *Abs. Surprise* is comparable across subject types. In Experiment B, there is no substantial difference in *Abs. Surprise* between subjects receiving correct or incorrect signals regarding TM1.

5 Results

Appendix C confirms that treatment assignment is well-balanced. We observe that the true productivity distributions are uniformly distributed (Figure C.1), while belief distributions deviate significantly from a uniform distribution (Figure C.2 and C.3). There are no significant differences between Individual Task Assignment (ITA) and Group Task Assignment (GTA) regarding the true productivity distributions and their belief distributions. These findings suggest that beliefs are biased, and the extent of these biases is comparable between both treatments.

We present our main results as follows. Section 5.1 shows that subjects adjust their beliefs with different objectives across treatments, consistent with Hypothesis 1. In Section 5.2, we present evidence supporting Hypothesis 2. We find that biased beliefs lead to greater inefficiency in Individual Task Assignment (ITA) than in Group Task Assignment (GTA). In Section 5.3, we show that subjects' beliefs consistently mismatch reality, not supporting Hypothesis 3.

5.1 Learning patterns

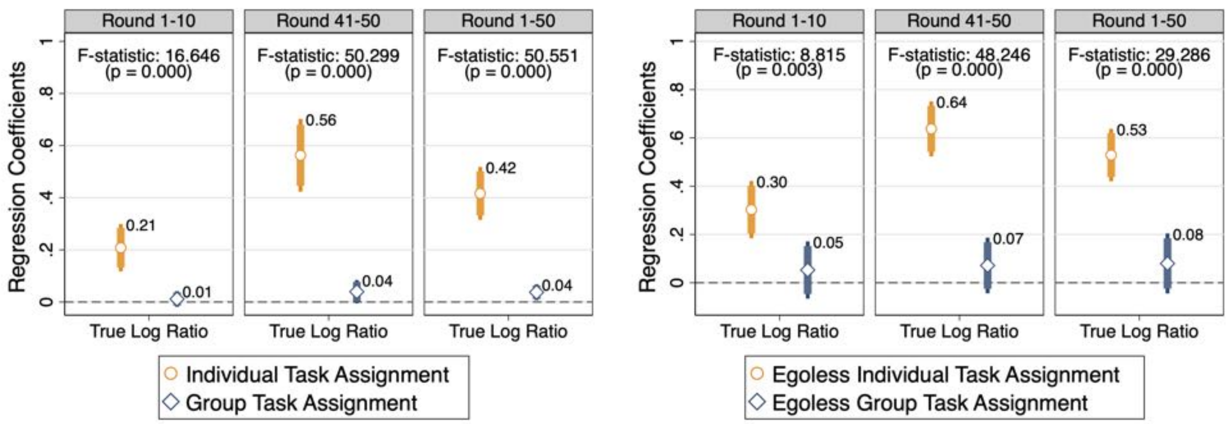
5.1.1 Learning the productivity ratio and the average productivity

To explore different learning patterns, as specified by Hypothesis 1, we regress subjects' beliefs (y_{it}) on the true value (y_i^*) as shown in Equation (3). i represents an individual and $t \in \{1, 2, \cdots, 50\}$ represents a round. As subjects have beliefs closer to the true value, the coefficient α_1 should approach 1. We control for rounds, and standard errors are clustered at the individual level.

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

First, we examine whether subjects learn about the productivity ratio between TM1 and

TM2 (Hypothesis 1.1). The *Log Ratio of Beliefs* is defined as the log ratio of beliefs about the productivity of TM1 to beliefs about the productivity of TM2. The *True Log Ratio* is the log ratio of the true productivity of TM1 and the true productivity of TM2. We regress the *Log Ratio of Beliefs* (y_{it}) on the *True Log Ratio* (y_i^*). Figure 3 plots the estimated α_1 of Equation (3).[17]



(a) Experiment A

(b) Experiment B

Figure 3: 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 H.1 and H.2 report the full regression estimates.

Figure 3a 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 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).

[17] Table H.1 and H.2 report the full regression estimates.

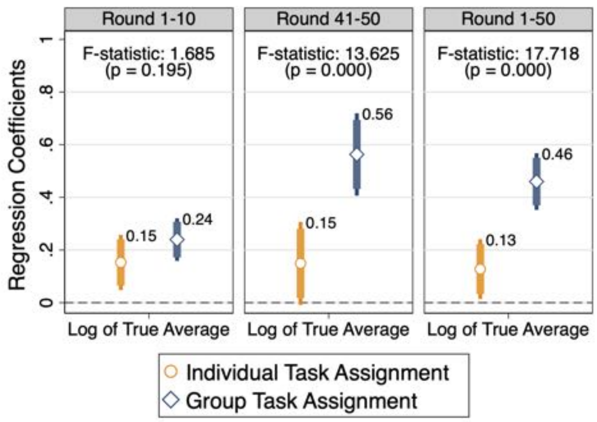
Figure 3b presents the results of Experiment B. The figure shows qualitatively identical learning patterns to those observed in Experiment A. Subjects in Egoless Individual Task Assignment (Egoless ITA) have beliefs closer to the productivity ratio between TM1 and TM2 more than subjects in Egoless Individual Group Assignment (Egoless GTA). The coefficient in Egoless ITA starts at around 0.3 in the first ten rounds and increases to 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).

Next, we examine whether subjects learn about the average productivity of TM1 and TM2 (Hypothesis 1.2). *Log of Average Beliefs* is defined by the log of beliefs about the average productivity of TM1 and TM2. *Log of True Average*, is the log of the true average productivity of TM1 and TM2. We regress *Log of Average Beliefs* (y_{it}) on *Log of True Average* (y_i^*).[18]

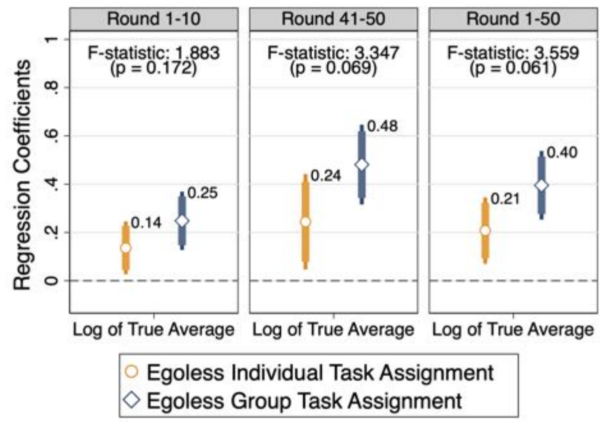
Figure 4a presents the results of Experiment A. The left and middle panels show the estimates during the first and last ten rounds, respectively. The right panel pools all 50 rounds. In ITA, the coefficient consistently remains below 0.2. On contrast, the coefficient in GTA is 0.24 in the first ten rounds, 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 4b 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

[18]Table H.3 and H.4 report the full regression estimates.



(a) Experiment A



(b) Experiment B

Figure 4: 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 H.3 and H.4 report the full regression estimates.

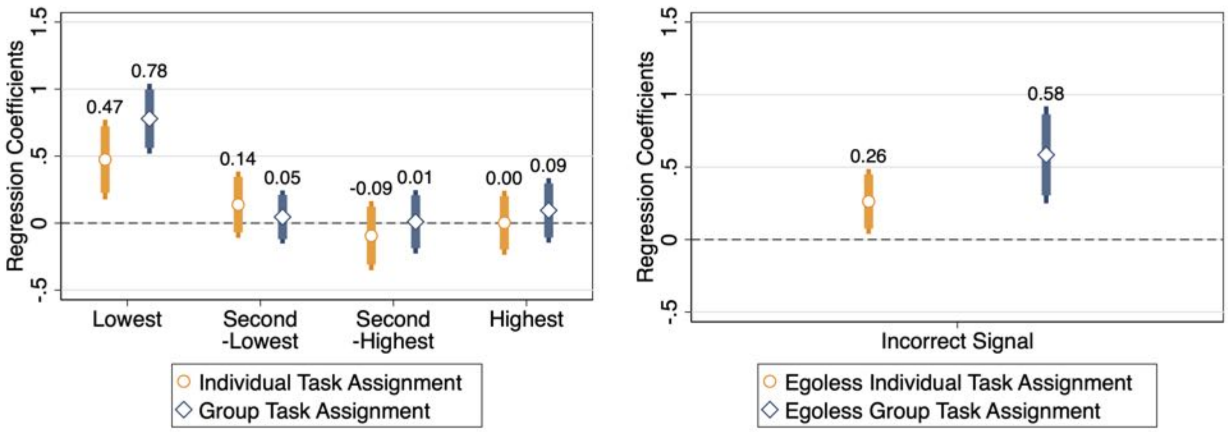
beliefs about the average (F-statistic: 1.883, p =0.172). However, the gap emerges over time. In the last ten rounds, belief biases among subjects in Egoless ITA is twice as large as those among subjects in Egoless GTA (F-statistic: 3.347, p =0.069). Pooling all rounds, the coefficient in Egoless GTA is significantly greater than that in Egoless ITA (F-statistic: 3.559, p =0.061).

5.1.2 Heterogeneous effects on learning

This subsection demonstrates heterogeneous learning effects. In Experiment A, we expect that subjects classified as *Lowest* and *Second-Lowest* types would hold more biased beliefs. In Experiment B, receiving incorrect signals about TM1 is expected to induce more biased beliefs. Our results substantiate the variations in belief biases across subjects as intended by our experiments.

We regress the absolute 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. In Experiment A, the coefficient β indicates the additional belief biases of a subject type in comparison to *Average* types (the omitted category). In Experiment B, the coefficient β captures the causal effect of receiving incorrect signals about TM1.

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



(a) Experiment A

(b) Experiment B

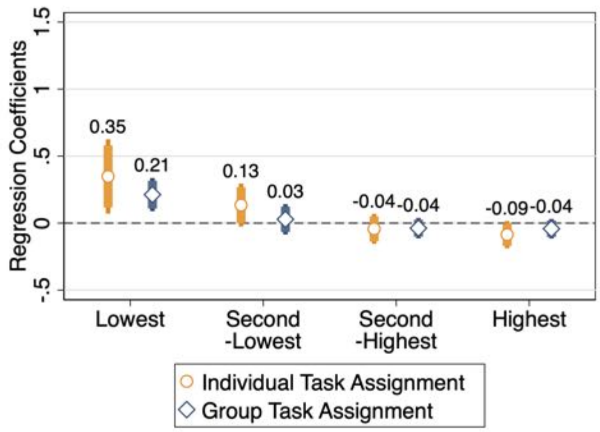
Figure 5: Biases in beliefs about the (true) productivity ratio

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 H.5 and H.6 report the full regression estimates.

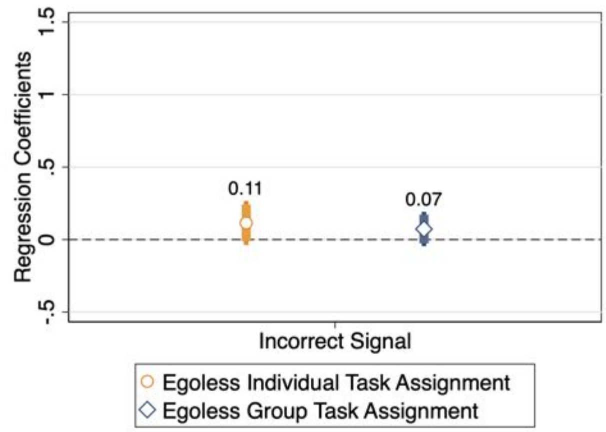
Figure 5 shows heterogeneous learning regarding the productivity ratio between TM1 and TM2.[19] Figure 5a shows that beliefs about the productivity ratio among *Lowest* types deviate further from the true ratio compared to *Average* types by 0.47 in Individual Task Assignment (ITA) and by 0.78 in Group Task Assignment (GTA). However, *Second-Lowest*,

[19] 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. Rounds are pooled together in both panels. We control for rounds, and standard errors are clustered at the individual level. Table H.5 and H.6 report the full regression estimates.

Second-Highest and *Highest* types do not exhibit more biased beliefs than *Average* types in both treatments. Figure 5b shows that receiving incorrect signals about TM1 increases beliefs biases by 0.26 in Egoless Individual Task Assignment (Egoless ITA) and by 0.58 in Egoless Group Task Assignment (Egoless GTA).[20]



(a) Experiment A



(b) Experiment B

Figure 6: 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 H.7 and H.8 report the full regression estimates.

Figure 6 shows heterogeneous learning regarding the average productivity of TM1 and TM2.[21] Figure 6a shows that in Individual Task Assignment (ITA) of Experiment A, belief biases of *Lowest* and *Second-Lowest* are greater than *Average* types' by 0.35 and 0.13, re-

[20]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 for those whose TM1 is a *Lowest* type. To account for potential impacts of TM1's type, Column (7)-(12) of Table H.6 report regression estimates controlling for indicators of TM1's types. The effect of receiving incorrect signals remains significant.

[21]The dependent variable 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. Rounds are pooled together in both panels. We control for rounds, and standard errors are clustered at the individual level. Table H.7 and H.8 report the full regression estimates.

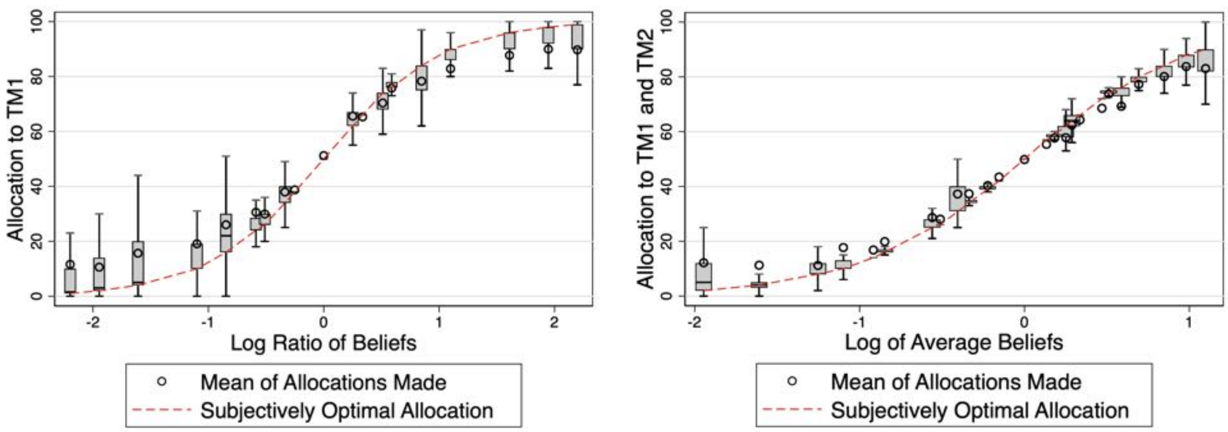
spectively. In Group Task Assignment (GTA), *Lowest* types hold more biased beliefs than *Average* types by 0.21. Note that the coefficients are smaller compared to coefficients found in Figure 5a). This suggests that learning about the average productivity of TM1 and TM2 is less affected by biased beliefs about TM1's productivity compared to learning about the productivity ratio between TM1 and TM2. This argument is strengthened by Figure 6b, providing causal evidence. Figure 6b shows that receiving correct signals in Experiment B does not significantly increase belief biases about the average productivity of TM1 and TM2. The causal effect of receiving an incorrect signal in Egoless Individual Task Assignment (Egoless ITA), 0.11, is not statistically different from zero (F-statistic: 2.24, $p = 0.137$). Similarly, the causal effect in Egoless Group Task Assignment (Egoless GTA), 0.07, is also not significantly different from zero (F-statistic: 1.49, $p = 0.225$). This is in contrast to the significant effect of receiving incorrect signals on beliefs about the productivity ratio between TM1 and TM2 (see Figure 5b).[22]

5.1.3 Allocation choices

This subsection confirms subjects' task allocation choices are consistent with their beliefs. Figure 7 shows box whisker plots of allocation choices made by subjects, pooling Experiment A and Experiment B.[23] Figure 7a shows allocation choices to TM1 in Individual Task Assignment (ITA) conditional on beliefs about the productivity ratio between TM1 and TM2. Figure 7b 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 the (perceived) expected output conditional on their beliefs.

[22] After controlling for potential impacts of TM1's type, the effect of receiving incorrect signals 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 H.8). Nonetheless, these coefficients are less than half of the coefficients for the productivity ratio between TM1 and TM2.

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



(a) Individual Task Assignment

(b) Group Task Assignment

Figure 7: 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.

The figures show that subjects make allocation choices that are predominantly consistent with subjectively optimal actions. We also find a strong linear relationship between subjects' allocation choices and their subjectively optimal choices (the coefficient: 0.831, F-statistic: 2607.42, p=0.000). Subject types are not correlated with the deviation between them. See Section D.

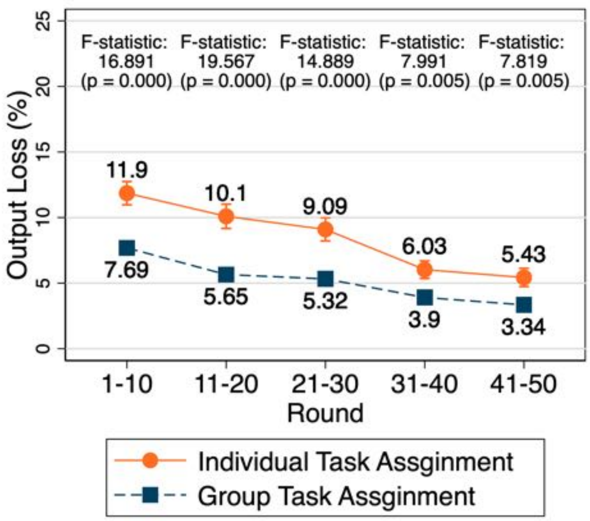
5.2 Treatment effects

This section tests Hypothesis 2. We show 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. This measure of allocative efficiency accounts for variations in potential output levels across treatment conditions and across the productivities of TM1 and TM2.[24]

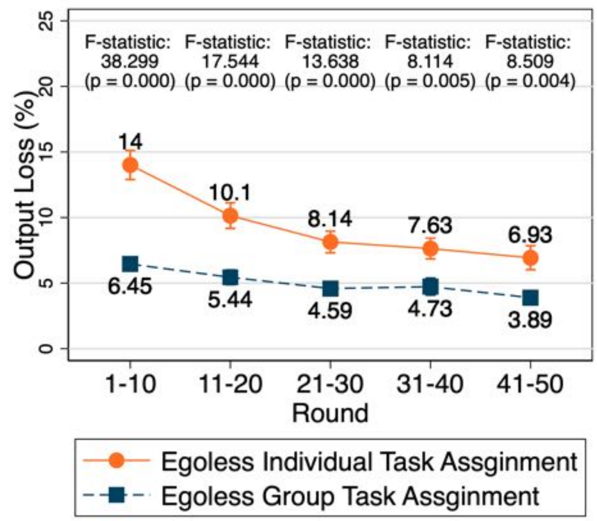
[24]Output loss is calculated as a percentage, ranging from 0 (indicating a subject achieving the highest possible output) to 100 (indicating achieving the lowest possible output). We divide the difference between

5.2.1 Treatment effects at the aggregate level

Figure 8 presents the average output loss by treatments. Figure 8a shows the results of Experiment A, and Figure 8b 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.



(a) Experiment A



(b) Experiment B

Figure 8: 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).

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 Exper-

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. 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.

iment 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. 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, skepticism may still be present concerning structural differences between Individual Task Assignment (ITA) and Group Task Assignment (GTA). To address this, Section 5.2.2 and 5.2.3 investigate heterogeneity in allocative efficiency within each treatment.[25]

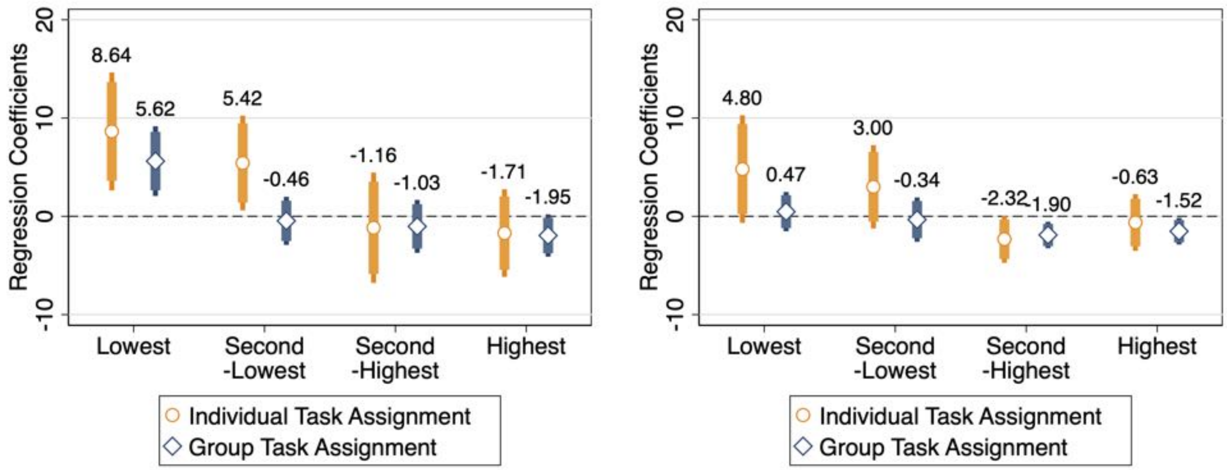
5.2.2 Treatment effects in Experiment A

Figure 9 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 additional percentage points of output loss incurred by a certain subject type on average compared to *Average* types. We control for rounds, and standard errors are clustered at the individual level.[26]

Figure 9a presents the results in 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,

[25] For instance, in Experiment A, *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 5.1.2. 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.

[26] Table H.11 reports the full regression estimates.



(a) Round 1-10

(b) Round 41-50

Figure 9: 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 H.11 report the full regression estimates.

in ITA, *Second-Lowest* types also experience higher output loss than *Average* types by 5.42 percentage points (p =0.028).

Figure 9b shows the results in 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 ($p = 0.64$).

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.[27] A joint test of the interactions being zero captures the treatment effects while fully controlling for structural differences between Individual Task Assignment (ITA) and Group Task Assignment (GTA). The result suggests a weak treatment effect. Pooling all rounds, we do not reject the null hypothesis that heterogeneous effects differ from ITA and GTA (F-statistic: 1.644, p=0.164).

In Experiment A, beliefs about TM1's productivity may be endogenous with subject types. Section E addresses this concern using the data from Experiment B. We find that Output Loss (%) in Experiment B is not correlated with either subjects' types or TM1's types.

5.2.3 Treatment effects in Experiment B

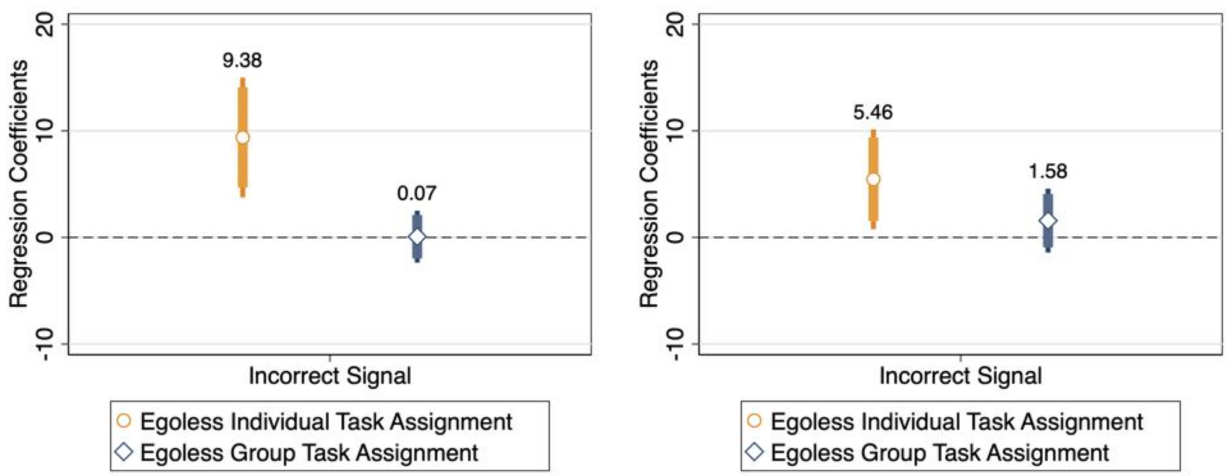
Figure 10 shows the causal effect of receiving incorrect signals about TM1 on allocative efficiency. We regress *Output loss (%)* on the indicator of receiving incorrect signals about TM1. The coefficients capture the difference in output loss between subjects who receive incorrect signals and those who receive correct signals (the omitted category). We control for rounds, and standard errors are clustered at the individual level.[28]

Figure 10a presents the results in 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 on Group Task Assignment (GTA) (p =0.001). Figure 10b presents the results in 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 remains insignificant.

To test the significance of the treatment effect, we regress on the indicators for receiving

[27] Table H.14 reports the regression estimates.

[28] Table H.12 reports the full regression estimates.



(a) Round 1-10

(b) Round 41-50

Figure 10: 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 indicator of receiving incorrect signals about TM1. The coefficients indicate the causal effect of receiving an incorrect signal about TM1 on output loss. We control for rounds, and standard errors are clustered at the individual level. Table H.12 reports the full regression estimates.

incorrect signals, treatment indicators, and their interactions.[29] A test of the interaction term being zero captures the treatment effects while fully controlling for structural differences between Individual Task Assignment (ITA) and Group Task Assignment (GTA). The result indicates that the effect of receiving incorrect signals significantly differs between treatments (F-statistic: 4.283, p=0.004).

5.2.4 Mechanisms

We consider three potential channels driving the treatment effects. (i) Biases in beliefs about the productivity ratio between TM1 and TM2. In Individual Task Assignment (ITA), making the optimal choice necessitates knowledge of the productivity ratio between TM1 and TM2. Therefore, we anticipate that *Output Loss (%)* in ITA is primarily driven by belief biases in the productivity ratio. (ii) Biases in beliefs about the average productivity

[29] Table H.15 reports the regression estimates.

of TM1 and TM2. The optimal choice requires knowledge of the average productivity of TM1 and TM2 in Group Task Assignment (GTA). We expect that biased beliefs about the average productivity drive a positive output loss in GTA. (iii) Mistakes in task allocation choices. Subjects may fail to optimize their allocation choices conditional on their beliefs. We anticipate that these choice mistakes have little impact.[30]

As shown in Section F, our analyses reveal that the different efficiency outcomes between Individual Task Assignment (ITA) and Group Task Assignment (GTA) stem from the distinct beliefs required to make an optimal choice in these treatments. Biased beliefs regarding the productivity ratio of TM1 and TM2 are a driving force towards inefficiency in ITA, while biased beliefs concerning the average productivity of TM1 and TM2 lead to inefficiency in GTA. Crucially, GTA offers protection against inefficiency resulting from belief biases, as learning the average productivity of TM1 and TM2 is less influenced by biased beliefs about TM1 than learning the productivity ratio of TM1 and TM2 (see Section 5.1.2).

5.3 Stable beliefs

To investigate Hypothesis 3.1, we compare the average *Abs. Surprise* in the first 10 rounds with that in the last ten rounds. Pooling Experiment A and Experiment B, 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). In both treatments, *Abs. Surprise* is significantly larger than zero in the last ten rounds (F-statistic: 282.46,

[30] In the analysis, we use absolute terms because the productivity space is bounded from *Lowest* to *Highest*. For example, biased 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. This specification aims to identify whether these potential errors in beliefs and choices have an impact on allocative efficiency and, if so, to what extent.

p=0.000 in ITA and F-statistic: 527.19, p=0.000).[31]

Next, we examine heterogeneity in *Abs. Surprise* (Hypothesis 3.2). For Experiment A, we regress *Abs. Surprise* on the indicators for subject types. A joint test of these indicators being zero captures whether there are significant differences in *Abs. Surprise* across subject types. In both Individual Task Assignment (ITA) and Group Task Assignment (GTA), we find significant heterogeneity in the first 10 rounds (F-statistic: 2.363, p=0.058 in ITA; F-statistic: 2.446, p=0.047 in GTA). However, the heterogeneity disappears in the last ten rounds (F-statistic: 1.471, p=0.217 in ITA; F-statistic: 1.103, p=0.356 in GTA). For Experiment B, we regress *Abs. Surprise* on the indicators for receiving incorrect signals. We find that subjects receiving incorrect signals experience significantly greater *Abs. Surprise* in ITA both in the first ten rounds (F-statistic: 4.44, p=0.038) and the last ten rounds (F-statistic: 4.039, p=0.047). However, the differences are not significant in GTA (F-statistic: 0.062, p=0.804 in the first ten rounds; F-statistic: 1.052, p=0.308 in the last ten rounds).[32]

6 Conclusions

We set out to study the circumstances under which biased beliefs have no consequences on optimal behavior. To do this, we design an environment where correct beliefs are not always necessary for optimal task allocation within a team. In particular, we exploited what Heidhues et al. (2018) calls *Non-Identifiability*, which states that biased beliefs are not detrimental if multiple unknown factors determining outputs can be summarized as a single statistic. To test this, we implemented a task assignment rule that incentivizes learning the average productivity of team members, as opposed to learning each team member's relative productivity. We find that the task assignment leads to more allocative efficiency despite the presence of biased beliefs about each team member's productivity.

[31] All standard errors are clustered at the individual level. See Figure H.3. Figure H.4 and H.5 present by Experiment A and Experiment B.

[32] All standard errors are clustered at the individual level. See Table H.16.

This study is relevant for organizations where belief biases are hard to change or identify. Our findings exemplify that biases in one dimension (e.g., one's competence) can prevent learning about other dimensions (e.g., others' competence). One behavioral bias might engender others. Importantly, we provide first evidence that some organization structures are robust to these behavioral biases. Our study highlights a fundamental identification problem intrinsic to equilibrium models with biased beliefs. Optimal behavior does not necessarily extrapolate across environments. Instead, the same biased beliefs can lead to optimal or sub-optimal behavior. Therefore, accounting for the environment's tolerance to behavioral biases is crucial to draw correct conclusions from observed behavior.

References

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

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2007), "Incentives for managers and inequality among workers: Evidence from a firm-level experiment." *The Quarterly Journal of Economics*, 122, 729–773.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), "Social connections and incentives in the workplace: Evidence from personnel data." *Econometrica*, 77, 1047–1094.

Bendor, J., A. Glazer, and T. Hammond (2001), "Theories of delegation." *Annual Review of Political Science*, 4, 235–269.

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.

Bertrand, Marianne and Sendhil Mullainathan (2001), "Are CEOs rewarded for luck? The ones without principals are." *The Quarterly Journal of Economics*, 116, 901–932.

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

Burgess, Simon, Carol Propper, Marisa Ratto, Stephanie von Hinke Kessler Scholder, and Emma Tominey (2010), "Smarter task assignment or greater effort: The impact of incentives on team performance." *The Economic Journal*, 120, 968–989.

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." *Working Paper*.

Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler (2022), "Aversion to Hiring Algorithms: Transparency, Gender Profiling, and Self-Confidence." *Working Paper*.

Delfgaauw, Josse, Robert Dur, and Michiel Souverijn (2020), "Team incentives, task assignment, and performance: A field experiment." *The Leadership Quarterly*, 31, 101241. Economics and Leadership.

Dlugosz, Jennifer, Yong Kyu Gam, Radhakrishnan Gopalan, and Janis Skrastins (Forthcoming), "Decision-making delegation in banks." *Management Science*.

Edelman, Benjamin and Michael Luca (2014), "Digital discrimination: The case of Airbnb.com." *Working Paper*.

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 (Forthcoming), "Mental models and learning: The case of base-rate neglect." *American Economic Review*.

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, 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." *Working Paper*.

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 (2022), "Inattentive inference." *Journal of the European Economic Asso-*

ciation, 21, 560–592.

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

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023), "Designing information provision experiments." *Journal of Economic Literature*, 61, 3–40.

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.

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

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

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

Huffman, David, Collin Raymond, and Julia Shvets (2022), "Persistent overconfidence and biased memory: Evidence from managers." *American Economic Review*, 112, 3141–75.

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), “How do expectations affect learning about fundamentals? some experimental evidence.” *Working Paper*.

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.

Neumark, David (2018), “Experimental research on labor market discrimination.” *Journal of Economic Literature*, 56, 799–866.

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: Current Research and Future Directions* (Emin Karagözoğlu and Kyle B. Hyndman, eds.), 455–475, Springer International Publishing, Cham.

Santos-Pinto, Luis and Leonidas Enrique de la Rosa (2020), “Overconfidence in labor markets.” In *Handbook of Labor, Human Resources and Population Economics* (Klaus F.

Zimmermann, ed.), 1–42, Springer International Publishing, Cham.

Sims, Christopher A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690.

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 H., Cass R. Sunstein, and John P. Balz (2014), "Choice architecture." In *The Behavioral Foundations of Public Policy* (Eldar Shafir, ed.).

Wolfers, Justin (2002), "Are voters rational? evidence from gubernatorial elections." *Working Paper*.

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.

Appendix

A 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

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

  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 $a_1 + a_2 = b_1 + b_2$.

B Pairs of stable beliefs and stable actions

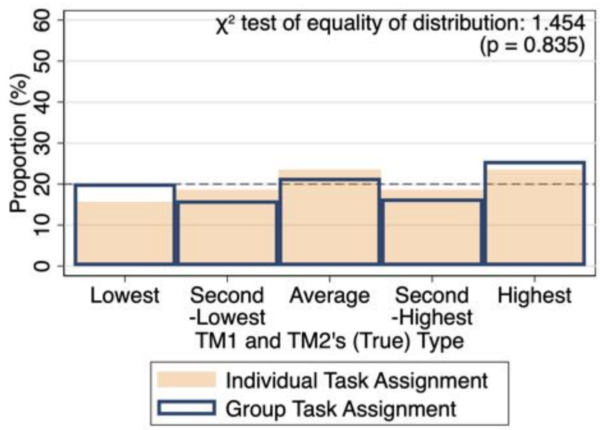
The table presents all possible pairs of stable, and biased, beliefs in Individual Task Assignment.

$(a_1,$	$a_2)$	$(b_1,$	$b_2)$	$x^*(b_1, b_2)$
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

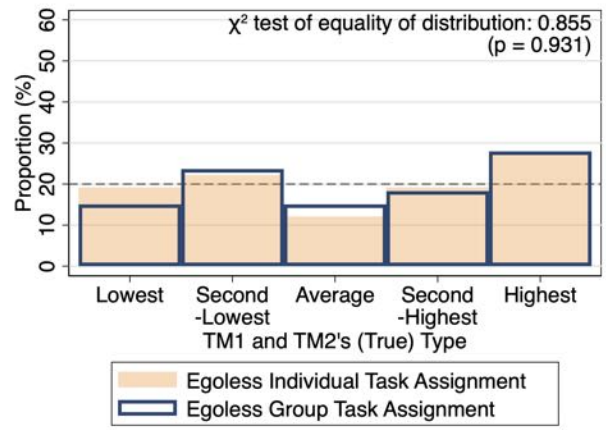
C Balanced treatment assignment

C.1 True productivity distributions

We test for differences in true productivity distributions between treatments.



(a) Experiment A



(b) Experiment B

Figure C.1: 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 C.1 shows that the treatments are balanced with respect to true productivity distributions. Figure C.1a 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 C.1b 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).

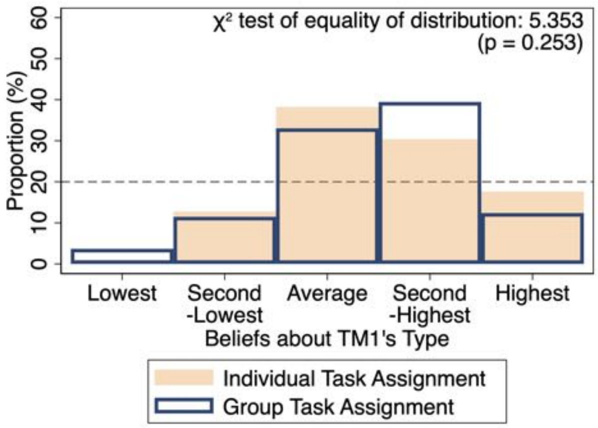
C.2 Belief distributions

Next, we analyze the distributions of beliefs about TM1 and TM2. Figure C.2 presents belief distributions in Experiment A. Figure C.3 presents belief distributions in Experiment B. Recall that in Experiment A, TM1 represents the subjects themselves whereas TM1 is a randomly-selected participant in Experiment B.

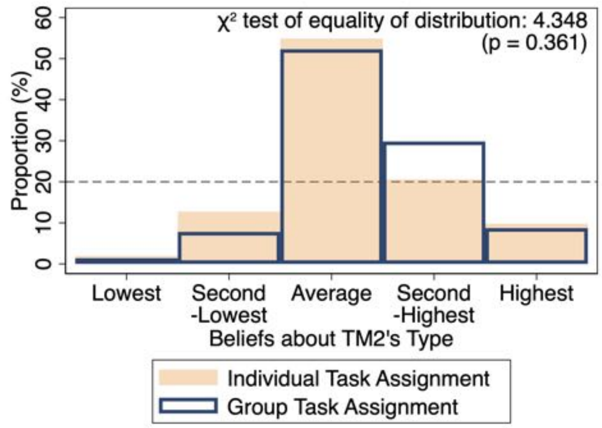
C.2.1 Experiment A

Figure C.2a 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, $p = 0.253$).

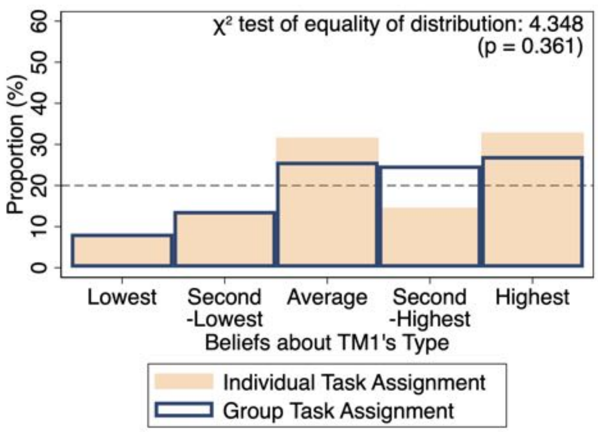
Figure C.2b presents the histograms of reported beliefs about TM2. Subjects also assign a slim chance to the teammate being the *Lowest* or the *Second-Lowest* in round 1. Most subjects initially report TM2's productivity as the *Average*. In Individual Task Assignment (ITA), 85.3% report that TM2's productivity is as high as the average, while in reality, it



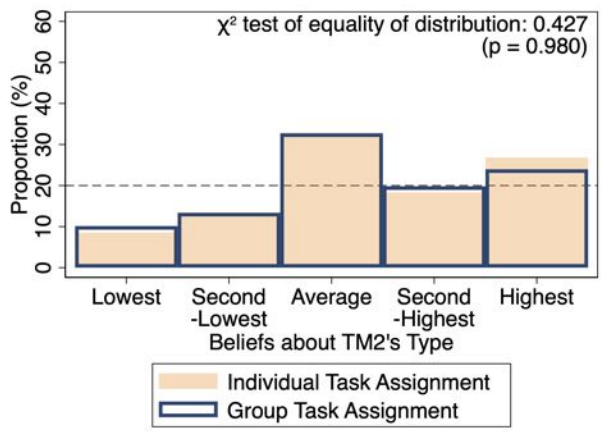
(a) Beliefs about TM1 (self) in Round 1



(b) Beliefs about TM2 in Round 1



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



(d) Beliefs about TM2 in Round 50

Figure C.2: 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).

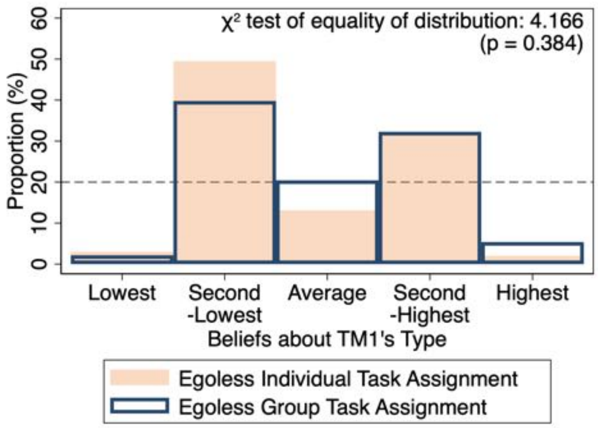
is 65.7%. In Group Task Assignment (GTA), 90.8% report that TM2's productivity is as high as the average, while in reality, it is 63.8%. The belief distributions are statistically different from the uniform distribution (Pearson χ^2 in ITA: 86.725, $p = 0.000$; Pearson χ^2 in GTA: 192.092, $p = 0.000$). They are not different across treatments (χ^2 test of equality of distribution: 4.348, $p = 0.361$).

Figure C.2c and C.2d 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 C.2a and C.2b). 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 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.

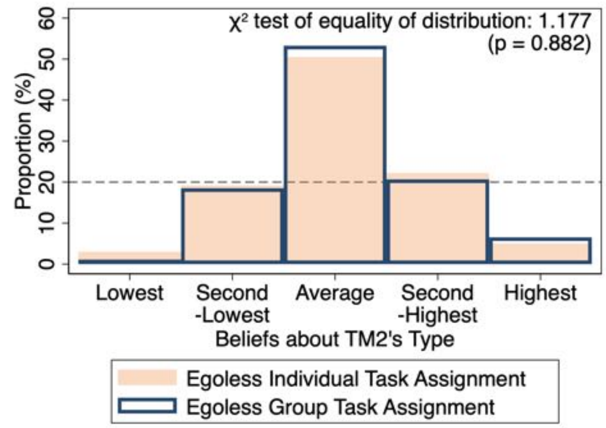
The changes in belief distributions from Round 1 to Round 50 are statistically significant. For TM1 (self), the χ^2 test of equality of distribution in Individual Task Assignment (ITA) is 14.531 ($p = 0.006$), while in Group Task Assignment (GTA), it results in 25.52 ($p = 0.000$). Regarding TM2, the 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, $p = 0.980$).

C.2.2 Experiment B

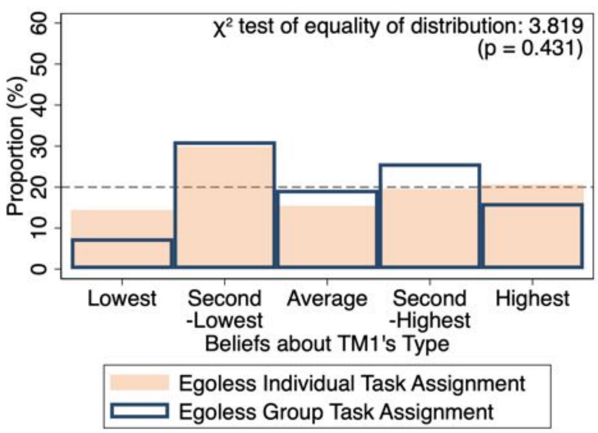
Figure C.3a and C.3b show the belief distributions of TM1' and TM2' productivities in the first round of Experiment B. When comparing these distributions to those in Experiment A, the overall shape suggests that when TM1 is a random person instead of the subject themselves, a greater proportion of subjects report the *Second-Lowest* productivity as a possibility. However, very few subjects report TM1 as the *Lowest*. The distributions are



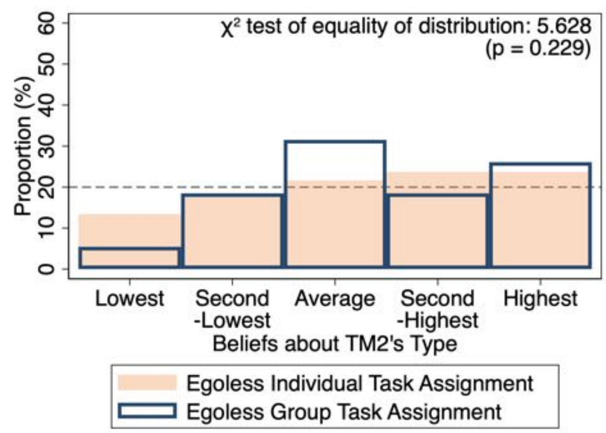
(a) Beliefs about TM1 in Round 1



(b) Beliefs about TM2 in Round 1



(c) Beliefs about TM1 in Round 50



(d) Beliefs about TM2 in Round 50

Figure C.3: 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).

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 C.3c and C.3d 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.

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.

D Subjectively optimal choices

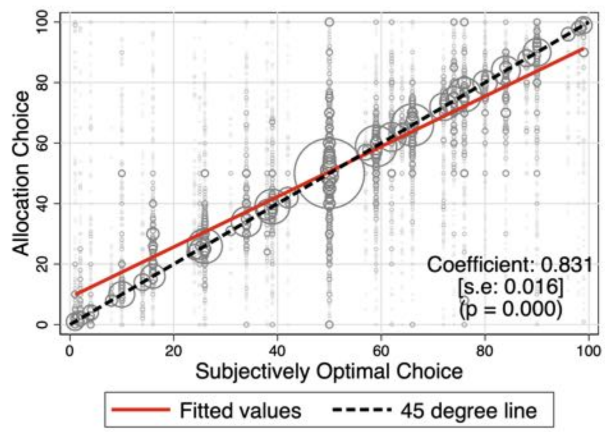


Figure D.1: Linear fit of allocation choice on subjectively optimal choice

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.

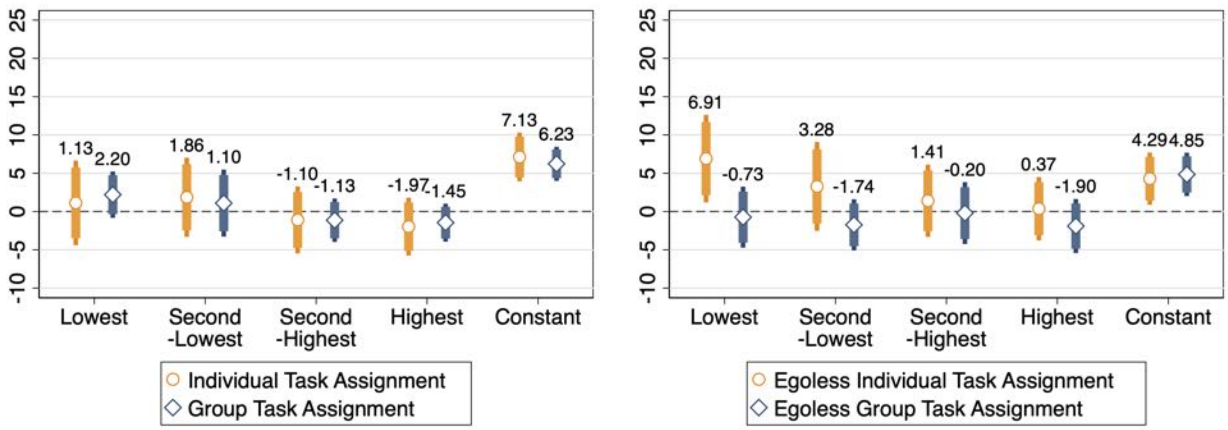
We regress choice allocations on the theoretical benchmark. If subjects subjectively optimize their allocation choices conditional on their beliefs, the coefficient should be 1. Standard errors are clustered at the individual level.

Figure D.1 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.[33] 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

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

to choose the optimal allocation based on their beliefs. The coefficient of the linear fit is 0.831 (F-statistic: 2607.42, $p = 0.000$).



(a) Experiment1

(b) Experiment B

Figure D.2: Absolute errors in choices by subject types

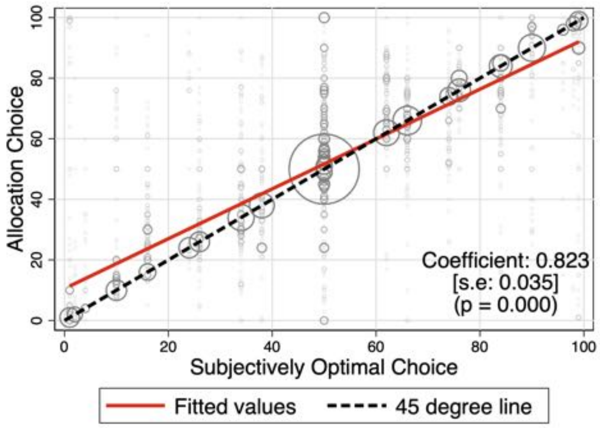
Note: The dependent variable is *Abs. Choice Errors*, defined by the absolute difference between the subjectively optimal action and their actual choice. *Lowest*, *Second-Lowest*, *Second-Highest*, and *Highest* refer to the indicators of subject types. The coefficients indicate the differences across 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 H.9 and H.10 report the full regression estimates.

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 avoid uncertainty regarding TM2's productivity (Benoît, Dubra, and Romagnoli, 2022).

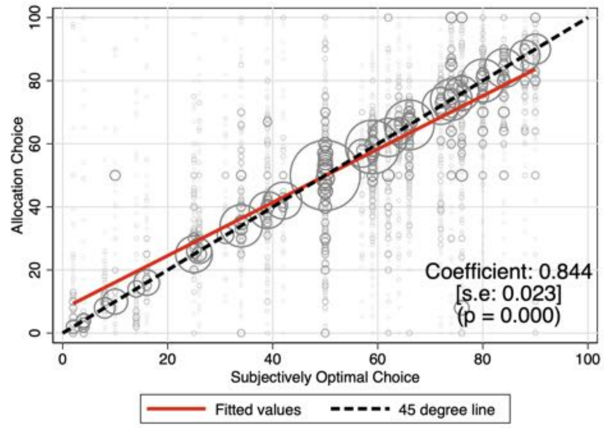
We regress *Abs. Choice Errors* on the indicators of subject types. *Abs. Choice Errors* is defined as the absolute difference between the subjectively optimal action and their actual choice. Figure D.2 confirms that there are no significant differences across subject types.[34] The exception is *Lowest* types in (Egoless) Individual Task Assignment of Experiment B. They deviate from their subjectively optimal choice by 7 hypothetical projects compared to

[34]Table H.9 and H.10 report the full regression estimates.

Average types (p=0.018).



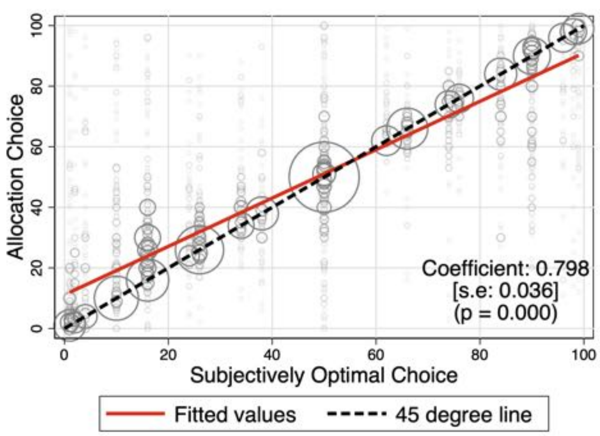
(a) Individual Task Assignment



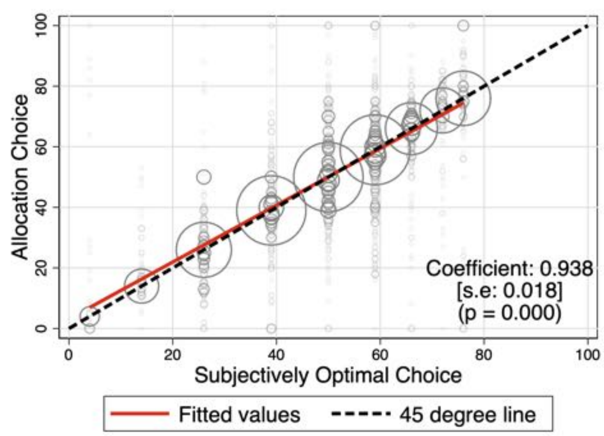
(b) Group Task Assignment

Figure D.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



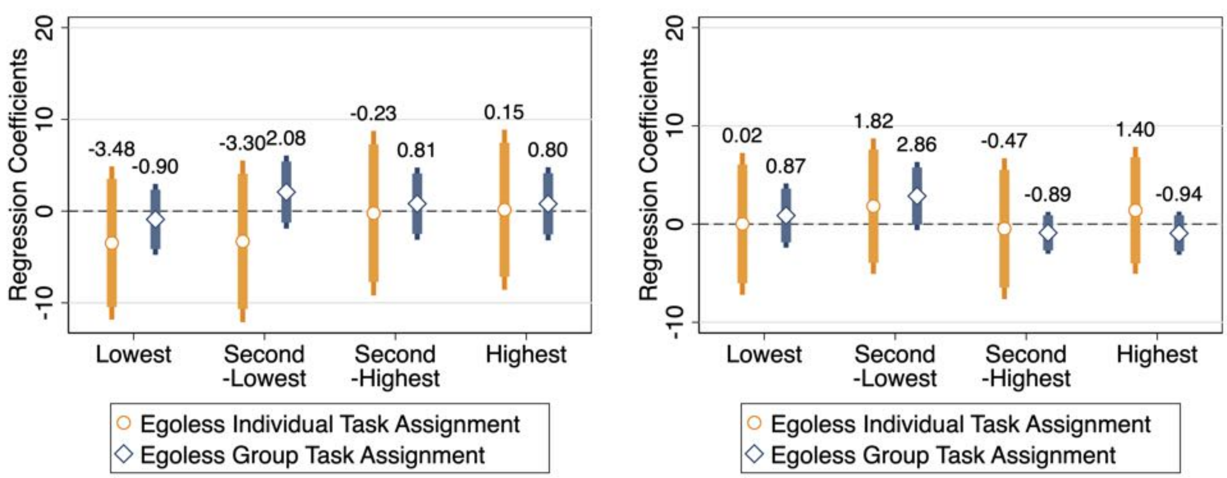
(b) Group Task Assignment

Figure D.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.

E Robustness checks

In Experiment A, TM1 represents the subject's self, and thus, the subject's type plays a crucial role. On the 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. Therefore, 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.



(a) Round 1-10

(b) Round 41-50

Figure E.1: 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 E.1 reports the full regression estimates.

Furthermore, we anticipate minimal influence from TM1's types on output loss in Exper-

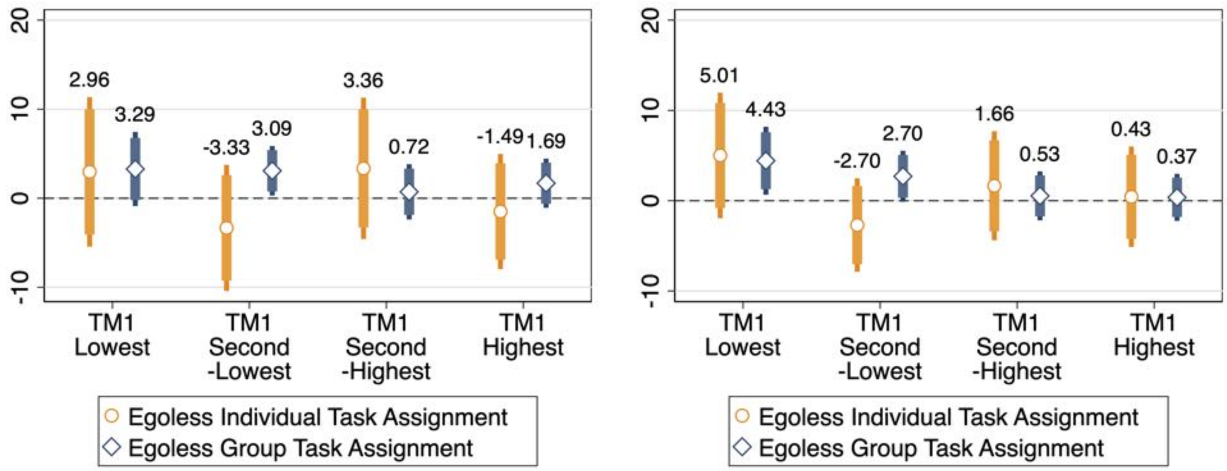
Table E.1: Allocative efficiency across subject types in Experiment B

	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 E.1.

iment B. Unlike subjects in Experiment A, especially those with below-average productivity, subjects in Experiment B are unlikely to 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.

We find no supporting evidence for these possibilities. Figure E.1 presents an identical figure to Figure 9, using data from Experiment B. Table E.1 report the full regression estimates. Ruling out the first concern, Figure E.1 demonstrates that output loss is not correlated with subject types in Experiment B. Figure E.2 presents the average output loss by



(a) Round 1-10

(b) Round 41-50

Figure E.2: 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 E.2 reports the full regression estimates.

TM1's type in Experiment B. Table E.2 report the full regression estimates. Addressing the second concern, the figure shows that the efficiency gaps between Individual Task Assignment (ITA) and Group Task Assignment (GTA), shown in Figure 8b, are not attributed to TM1's types.

Table E.2: 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 E.2.

F Mechanisms

We regress *Output Loss (%)* on three regressors. (i) *Abs. Errors in Log Ratio on Beliefs*. It represents 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*. It represents 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*. It represents the absolute difference between the (subjectively) optimal allocation choice and the actual choice.

Table F.1 and F.2 report the regression estimates for Individual Task Assignment (ITA) and Group Task Assignment (GTA), respectively. 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.

In Table F.1, 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 9, *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 H.11.) However, these effects vanish after accounting for the biases (the coefficient

Table F.1: 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.

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 about TM1 increases output loss by 6.082 percentage points ($p = 0.003$). (See Column (3) of Table H.12.) However, after accounting for belief biases, the causal effect reduces to 2.257 (p =0.078).

Table F.2 highlights beliefs about the average productivity of TM1 and TM2 play a significant role 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 H.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 H.12.) After

Table F.2: 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.

accounting for belief biases, the causal effect drops further to 0.855 ($p = 0.269$).

G Additional figures and tables

Table H.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 3a.

Table H.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 3b.

Table H.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 4a.

Table H.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 4b.

Table H.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 5a.

Table H.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 5b.

Table H.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 : Lowest$_{ITA}$ = Lowest$_{GTA}$						
F-statistic		0.089		1.052		0.787
p-value		0.765		0.306		0.376
H_0 : Second-Lowest$_{ITA}$ = Second-Lowest$_{GTA}$						
F-statistic		1.600		0.390		1.159
p-value		0.207		0.533		0.282
H_0 : Second-Highest$_{ITA}$ = Second-Highest$_{GTA}$						
F-statistic		0.085		0.029		0.006
p-value		0.771		0.865		0.939
H_0 : Highest$_{ITA}$ = 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 6a.

Table H.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 6b.

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

	DV: Abs. Errors in Choices					
	Round 1-10		Round 41-50		Round 1-50	
	(1)	(2)	(3)	(4)	(5)	(6)
	ITA	GTA	ITA	GTA	ITA	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 D.2a.

Table H.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 D.2b.

Table H.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 9.

Table H.12: Treatment effects in Experiment B

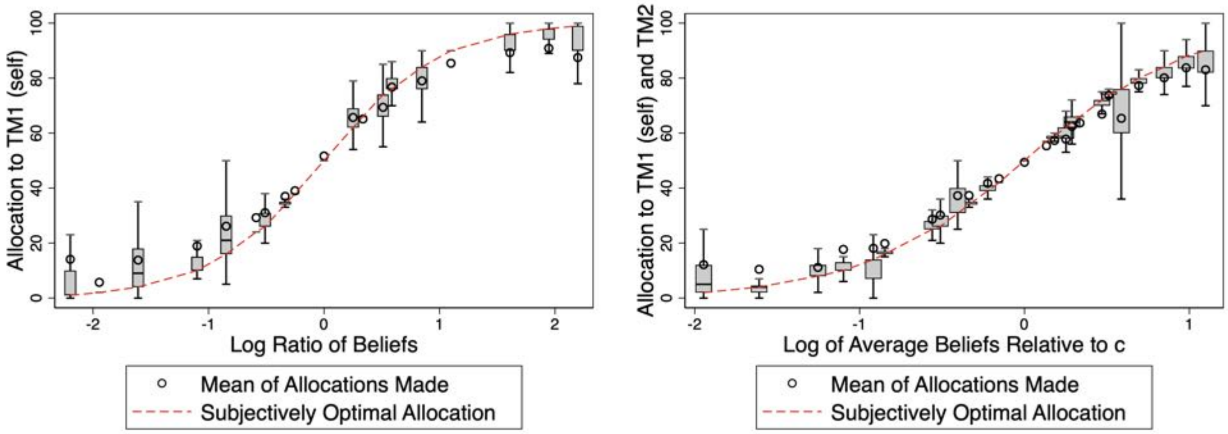
	DV: Output loss (%)					
	Individual Task Assignment			Group Task Assignment		
	(1)	(2)	(3)	(4)	(5)	(6)
	Round 1-10	Round 41-50	Round 1-50	Round 1-10	Round 41-50	Round 1-50
Incorrect Signal	9.383***	5.457**	6.082***	0.074	1.577	1.248
	(2.831)	(2.359)	(2.026)	(1.227)	(1.509)	(1.174)
Round	-0.280	0.176	-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 10.

Table H.13: Experiment A by different values of Robot's productivity

	DV: Output loss (%)	
	(1)	(2)
Robot's productivity = 30	-0.866	-0.892
	(0.684)	(1.393)
Robot's productivity = 70	0.850	-1.768*
	(0.795)	(1.016)
Lowest		0.330
		(1.348)
Lowest × Robot's productivity = 30		3.412
		(2.504)
Lowest × Robot's productivity = 70		5.090**
		(2.516)
Second-Lowest		0.758
		(1.406)
Second-Lowest × Robot's productivity = 30		-3.034
		(1.935)
Second-Lowest × Robot's productivity = 70		1.578
		(2.343)
Second-Highest		-1.086
		(0.953)
Second-Highest × Robot's productivity = 30		-0.668
		(1.817)
Second-Highest × Robot's productivity = 70		1.952
		(1.673)
Highest		-1.559
		(1.023)
Highest × Robot's productivity = 30		0.006
		(1.649)
Highest × Robot's productivity = 70		3.613**
		(1.626)
Round	-0.105***	-0.105***
	(0.011)	(0.011)
Constant	7.891***	8.282***
	(0.512)	(0.843)
R^2	0.021	0.042
Total Observations	10900	10900
Num. of Individuals	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). We control for rounds, and standard errors are clustered at the individual level are in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance.

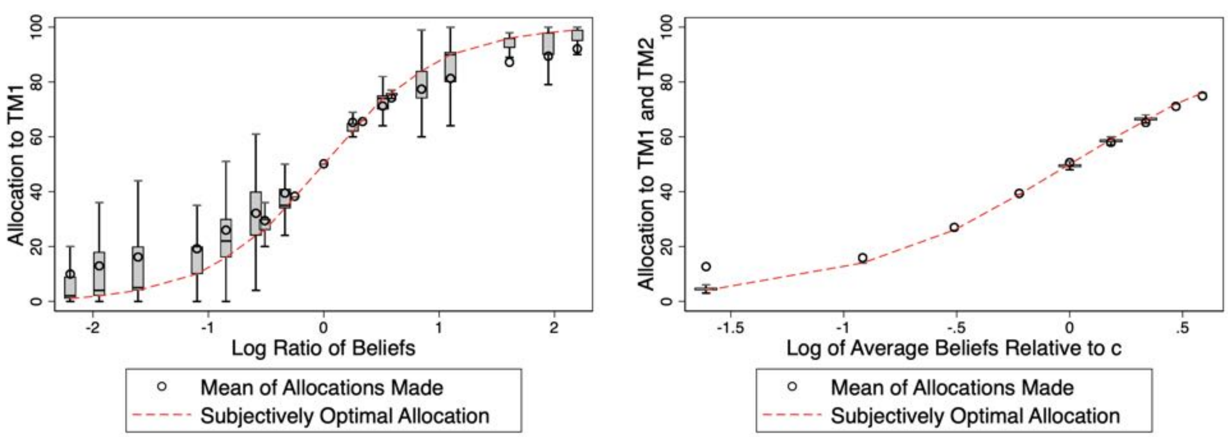


(a) Individual Task Assignment

(b) Group Task Assignment

Figure H.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 H.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.

Table H.14: Treatment effects in Experiment A (fully saturated)

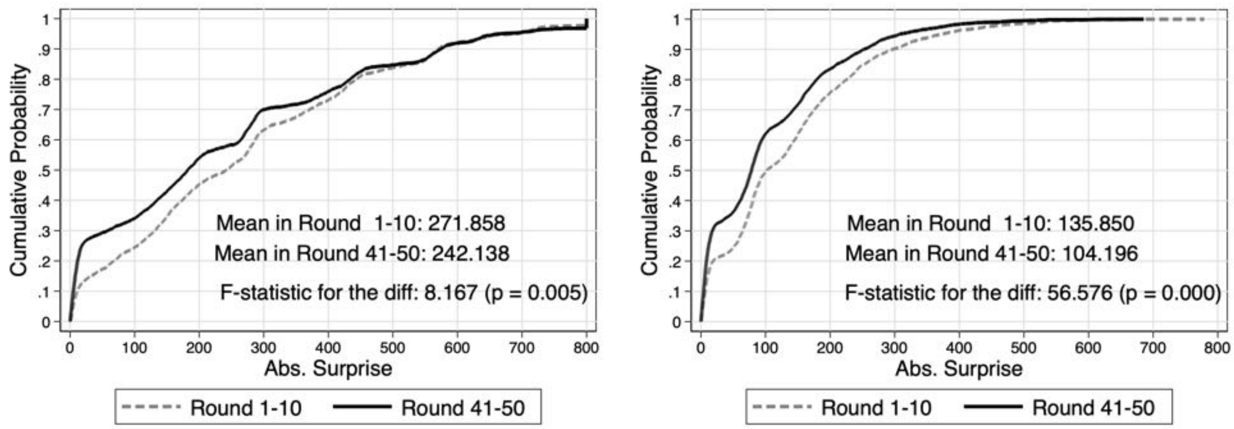
	DV: Output Loss (%)		
	Round 1-10	Round 41-50	Round 1-50
	(1)	(2)	(3)
Lowest	8.636***	4.806*	6.354***
	(3.014)	(2.749)	(2.180)
Second-Lowest	5.424**	3.004	3.271*
	(2.419)	(2.127)	(1.889)
Second-Highest	-1.164	-2.336*	-1.464
	(2.820)	(1.208)	(1.510)
Highest	-1.707	-0.625	-1.135
	(2.238)	(1.450)	(1.567)
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 H.15: 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$\times GTA$	-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.

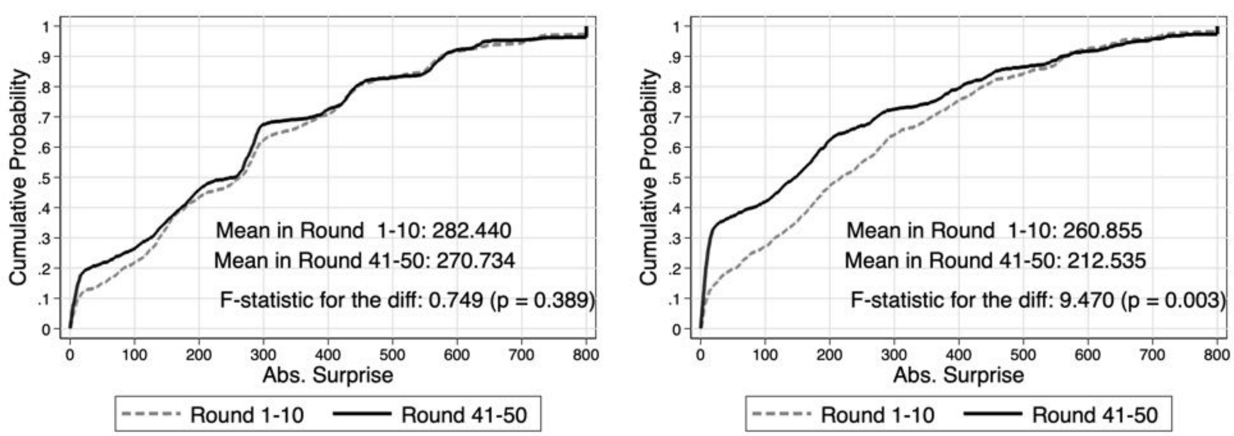


(a) Individual Task Assignment

(b) Group Task Assignment

Figure H.3: 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. Standard errors are clustered at the individual level.

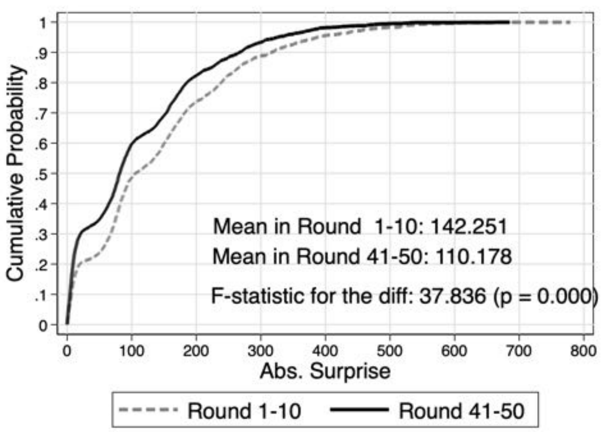


(a) Experiment A

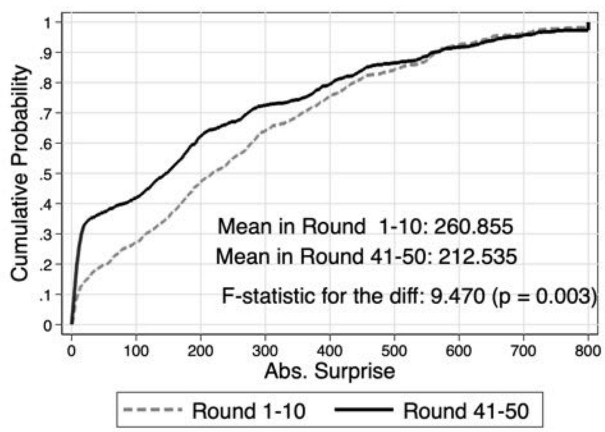
(b) Experiment B

Figure H.4: Cumulative distributions of *Abs. Surprise* for Individual Task Assignment

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. Values exceeding 800 are capped at 800. Standard errors are clustered at the individual level.



(a) Experiment A



(b) Experiment B

Figure H.5: Cumulative distributions of *Abs. Surprise* for Group Task Assignment

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. Values exceeding 800 are capped at 800. Standard errors are clustered at the individual level.

Table H.16: Heterogeneity in *Abs. Surprise*

	DV: Abs. Surprise							
	Experiment A				Experiment B			
	ITA		GTA		Egoless ITA		Egoless GTA	
	(1) Round 1-10	(2) Round 41-50	(3) Round 1-10	(4) Round 41-50	(5) Round 1-10	(6) Round 41-50	(7) Round 1-10	(8) Round 41-50
Lowest	129.897**	142.998*	49.842***	7.861				
	(60.332)	(77.691)	(18.970)	(19.021)				
Second-Lowest	76.619	8.014	-5.845	-23.066				
	(47.187)	(60.099)	(16.977)	(16.061)				
Second-Highest	6.445	-21.358	-1.935	-4.714				
	(40.963)	(41.889)	(15.590)	(16.919)				
Highest	-25.504	-39.933	2.862	6.507				
	(36.363)	(43.348)	(14.560)	(16.567)				
Incorrect Signal					86.685**	96.110**	4.660	20.150
					(41.139)	(47.823)	(18.743)	(19.644)
Constant	252.569***	259.993***	132.729***	111.410***	237.769***	186.866***	119.529***	85.401***
	(24.848)	(30.274)	(11.011)	(11.509)	(17.277)	(23.083)	(7.802)	(6.998)
Test for Heterogeneity								
F-statistic	2.363	1.471	2.446	1.103	4.440	4.039	0.062	1.052
p-value	0.058	0.217	0.047	0.356	0.038	0.047	0.804	0.308
R^2	0.064	0.067	0.027	0.009	0.032	0.032	0.000	0.008
Total Observations	1019	1000	2178	2178	980	966	917	929
Num. of Individuals	102	102	218	218	99	97	92	93

Note: The dependent variable is *Abs. Surprise*. For Experiment A, we regress on the indicators for subject types. For Experiment B, we regress on the indicators for receiving incorrect signals. The reported F-statistics and the corresponding p-values test if a joint test of the indicators being zero.

H Study materials

We present the screens encountered by subjects during the experiments. Part 1 is common to all treatments. Screenshots for Part 2 are ordered as follows: Individual Task Assignment in Experiment A, Group Task Assignment in Experiment A, Individual Task Assignment in Experiment B, and Group Task Assignment in Experiment B.

Time left to complete this page: **1:52**

PART 1

Sliders 1-10

Sliders 11-20

Sliders 21-30

Sliders 31-40

Sliders 41-50

Sliders 51-60

Sliders 61-70

Sliders 71-80

Sliders 81-90

Sliders 91-100

PART 2

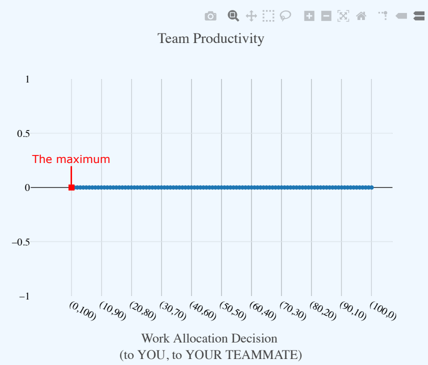
Round 1 (out of 50)

Task A: Guess

YOUR Individual Productivity	--- Select---
YOUR TEAMMATE's Individual Productivity	--- Select---

Calculate

Simulator



Conditional on that your individual productivity is **0** and your teammate's individual productivity is **0**, the team productivity is maximized when you allocate **0** to YOU and **100** to YOUR TEAMMATE.

The team is expected to produce **0.00** points.

Task B: Work Allocation Decision

NaN to YOU | **NaN** to YOUR TEAMMATE

History

Round	Final Guess about Individual Productivity		Work Allocation Decision		Points You Earned (= Team Productivity + A Lucky Number)
	You	Your Teammate	You	Your Teammate	

Submit the Decisions

Task A	
YOUR Individual Productivity	**0**
YOUR TEAMMATE's Individual Productivity	**0**
Task B	
Work Allocation to YOU	**NaN**
Work Allocation to YOUR TEAMMATE	**NaN**
I confirm that this is my final decisions in this round.	☐

Submit

PART 2

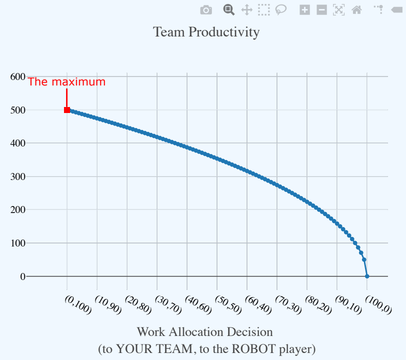
Round 1 (out of 50)

Task A: Guess

YOUR Individual Productivity	--- Select---
YOUR TEAMMATE's Individual Productivity	--- Select---

Calculate

Simulator



Conditional on that your individual productivity is **0** and your teammate's individual productivity is **0**, the team productivity is maximized when you allocate **0** to YOUR TEAM and **100** to the ROBOT player. The team is expected to produce **500.00** points.

Task B: Work Allocation Decision

NaN to YOUR TEAM **NaN** to the ROBOT player

History

Round	Final Guess about Individual Productivity		Work Allocation Decision		Points You Earned (= Team Productivity + A Lucky Number)
	You	Your Teammate	Your Team (You & Your Teammate)	Robot player	

Submit the Decisions

Task A	
YOUR Individual Productivity	0
YOUR TEAMMATE's Individual Productivity	0
Task B	
Work Allocation to YOUR TEAM	NaN
Work Allocation to the ROBOT player	NaN
I confirm that this is my final decisions in this round.	☐

Submit

PART 2

Round 1 (out of 50)

Task A: Guess

Teammate 1's Individual Productivity	--- Select---
Teammate 2's Individual Productivity	--- Select---

✔ The following message is true with a 75 percent chance:

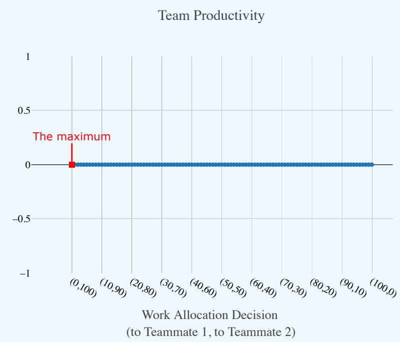
Individual productivity of Teammate 1 is below 50.

Calculate

History

Round	Final Guess about Individual Productivity		Work Allocation Decision		Points You Earned (= Team Productivity + A Lucky Number)
	Teammate 1	Teammate 2	Teammate 1	Teammate 2	

Simulator



Conditional on that Teammate 1's individual productivity is **0** and Teammate 2's individual productivity is **0**, the team productivity is maximized when you allocate **0** to Teammate 1 and **100** to Teammate 2.

The team is expected to produce **0.00** points.

Task B: Work Allocation Decision

NaN to Teammate 1 | **NaN** to Teammate 2

Submit the Decisions

Task A	
Teammate 1's Individual Productivity	**0**
Teammate 2's Individual Productivity	**0**
Task B	
Work Allocation to Teammate 1	**NaN**
Work Allocation to Teammate 2	**NaN**
I confirm that this is my final decisions in this round.	☐

Submit

PART 2

Round 1 (out of 50)

Task A: Guess

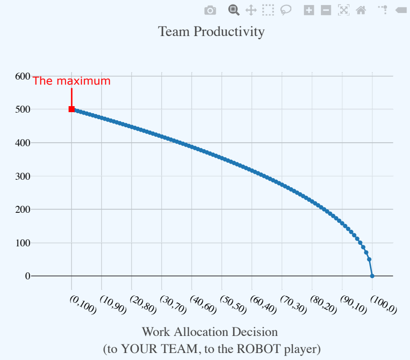
Teammate 1's Individual Productivity	--- Select---
Teammate 2's Individual Productivity	--- Select---

✔ The following message is true with a 75 percent chance:

Individual productivity of Teammate 1 is above or equal to 50.

Calculate

Simulator



Conditional on that Teammate 1's individual productivity is **0** and Teammate 2's individual productivity is **0**, the team productivity is maximized when you allocate **0** to YOUR TEAM and **100** to the ROBOT player. The team is expected to produce **500.00** points.

Task B: Work Allocation Decision

NaN to YOUR TEAM

NaN to the ROBOT player

History

Round	Final Guess about Individual Productivity		Work Allocation Decision		Points You Earned (= Team Productivity + A Lucky Number)
	Teammate 1	Teammate 2	Your Team (Teammate 1 & Teammate 2)	Robot player	

Submit the Decisions

Task A	
Teammate 1's Individual Productivity	**0**
Teammate 2's Individual Productivity	**0**
Task B	
Work Allocation to YOUR TEAM	**NaN**
Work Allocation to the ROBOT player	**NaN**
I confirm that this is my final decisions in this round.	☐

Submit