Motivating Beliefs in a Just World

Click here for the most recent version.

Max Lobeck*

June 15, 2021

Abstract

This paper studies whether individuals distort their beliefs about the relative importance of effort and luck to motivate themselves to exert effort. To that end, I develop a novel experimental design where past experience of success or failure serves as a noisy signal about the true importance of effort in achieving success. To test whether individuals distort their beliefs to motivate future effort, I vary the moment in time when subjects are informed about an effortful task to be performed later in the experiment. Subjects who receive the information before belief elicitation face an incentive to distort their beliefs to motivate effort in the later task. The results show that such individuals are more likely to believe that their effort is important for success. Motivating belief distortion is particularly pronounced for subjects who receive disincentivizing news about the true state of the world, i.e. that success depends on luck rather than on effort. I additionally test whether motivating belief distortion affects subjects' willingness to distribute money between two *other* individuals as a third-party spectator. I find no evidence that distributive behavior differs across the two treatment groups. These results suggest that individuals' luck-effort beliefs not only depend on past or current events but are also endogenous to the incentive structure individuals expect to face.

JEL: C91, D83, D63

Keywords: Motivated Beliefs, Fairness, Online Laboratory Experiment, Redistribution

*Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, max.lobeck@psemail.eu. I thank Nicolas Jacquemet, Claudia Senik, Kai Barron, Beatrice Boulu-Reshef, Alessandra Casella, Sophie Cetre, Liza Charroin, Urs Fischbacher, Marc Fleurbaey, Florian Grosset, Jeanne Hagenbach, Nina Hestermann, Shachar Kariv, Fabrice Le Lec, Suanna Oh, Leonardo Pejsachowicz, Stefano Piasenti, Chris Roth, Morten Støstad, Charlotte Saucet, Robert Stüber, Michael Thaler, Bertil Tungodden, Nikhil Vellodi, Florian Zimmermann, as well as seminar and conference participants at Aarhus University, Berlin School of Economics, NHH Bergen, Paris School of Economics and conference participants at the RGS Doctoral Conference, ADRES Doctoral Conference, MBEES and the HEIRS Workshop "Illustion of Merit", for very helpful comments and discussions. Funding by CEPREMAP (Program 3: Public economics and redistribution) and EUR grant ANR-17-EURE-0001 is gratefully acknowledged. This experiment was pre-registered (AEARCTR-0006421) and it was approved by PSE's IRB (2020-15).

1 Introduction

Research in recent decades has made great strides in identifying factors that shape attitudes towards redistribution. Such factors include beliefs about the source of inequality and the extent to which income is determined by luck rather than effort (luck-effort beliefs) (Alesina and Giuliano, 2011; Almås et al., 2020). While the literature has recognized the importance of these beliefs, empirical evidence about how they are formed remains limited.

Beliefs about the relative importance of luck and effort may be motivated. If beliefs are motivated, individuals distort their beliefs about the true relationship between effort and success in order to fulfill a certain goal. For example, if individuals have a strong desire to maintain a positive self-image by believing that they are more talented and productive than their peers, they want to attribute failure to luck and success to effort (Deffains et al., 2016). Similarly, if individuals have a strong desire to overcome a lack of motivation, they have an incentive to maintain the view that effort is important for success (Bénabou and Tirole, 2006).

This paper studies whether individuals distort luck-effort beliefs to motivate effort. I call this form of motivated beliefs motivat*ing* beliefs. Bénabou and Tirole (2002) propose that individuals distort beliefs to counter an under-provision of effort due to self-control problems. These same authors created a model that shows how motivating beliefs affect preferences for redistribution (Bénabou and Tirole, 2006). Specifically, they showed that, if an economic agent with self-control problems expects low levels of redistribution, exerting low levels of effort can become very costly. This creates a demand to motivate future effort by distorting beliefs. To validate this model, it is important to provide empirical evidence of motivating belief distortion, because such evidence would show that luck-effort beliefs are shaped by expectations about future levels of redistribution. This evidence would advance our understanding of the dynamic interaction between inequality, redistributive preferences, and beliefs about the importance of luck and effort. Such evidence also would imply that the causal relationship between luck-effort beliefs and redistribution runs both ways: Beliefs affect the demand for redistribution and expected levels of redistribution affect beliefs by shaping incentives.[1]

To test the prediction that future incentives distort beliefs about the importance of effort in achieving success I use an online experiment. In the experiment, subjects begin by performing a cumbersome real effort task. This task is completed in an uncertain environment, where the payment rule depends on the state of the world (Environment) that was drawn at the beginning of the experiment. If the subject is in the Performance-Environment condition, the likelihood of winning a prize for completing the task is an increasing function of their performance on the task. Specifically, the subject participates in a noisy tournament against a randomly drawn competitor, where the chance of winning the prize is equal to 80% if the subject transcribes more

[1] Beliefs affect preferences for redistribution in several ways: First, there are selfish reasons to ask for less redistribution if one distorts beliefs to motivate future effort. An individual who believes that she will be a net contributor if her effort is reflected in the pre-tax income distribution would be less likely to support redistribution after engaging in motivating belief distortion. Second, motivating belief distortion may affect preferences for redistribution for other-regarding reasons. Meritocratic individuals who distort their beliefs in a motivating way are less likely to believe that an initially unequal distribution is due to luck and if they accept inequalities that reflect differences in effort, they should, hence, opt for less redistribution compared to a situation where they did not distort their beliefs to motivate future effort.

images than her competitor, while the chance is equal to 20% if she transcribes fewer images. If the subject is in the Chance-Environment condition, her performance on the task has no effect on her likelihood of winning the prize; the subject wins the prize with 50% probability no matter how many words she transcribes.

Upon completing the task, subjects receive a noisy signal informing them about which condition—Performance-Environment or the Chance-Environment—they were in. The aim of this signal is to induce variation in baseline beliefs regarding the state of the world. I induce these beliefs by giving subjects feedback about the outcome of the first task. The feedback comprises two pieces of information: (1) whether a subject won the prize and (2) whether she transcribed more or fewer images than her competitor. Using this information, subjects can form posterior beliefs about the likelihood of being in the Chance- or Performance-Environment. For example, a subject who learns that she transcribed more images than the competitor but did not win the prize is likely to perceive herself as having a high probability of being in the Chance-Environment, that is, the condition in which success is unrelated to effort. By contrast, if the same person learns that she won the prize, she should perceive herself as having a high probability of being in the Performance-Environment where effort does influence likelihood of success.[2] After subjects receive the signal, I elicit their probabilistic beliefs about the environment (Chance or Performance).

To test whether subjects distort baseline beliefs to motivate themselves to exert effort, I introduce a second task that subjects can complete at the end of the experiment and that serves as an incentive (and, hence, motive) to distort beliefs for motivating purposes. As in the first task, the payment rule depends on the environment that was drawn at the beginning of the experiment: If the subject is in the Performance-Environment, effort determines whether one receives a reward for performance on the second task, while for subjects in the Chance-Environment, effort has no effect on the likelihood of receiving a reward. Subjects who know about the second task may overestimate their likelihood of being in the Performance-Environment in order to motivate themselves to work hard on the second task.

To identify motivating belief distortion, I vary the point in time at which I inform subjects about the second task: Subjects who are assigned to the Motive treatment group are informed about the second task *before* belief elicitation and, hence, have an incentive to distort beliefs to motivate effort. Subjects in the No-Motive treatment group receive this information *after* belief elicitation. The latter subjects have no incentive to distort beliefs to motivate effort because they do not know that they will be completing a second task in the experiment. This variation allows me to test the main hypothesis of the experiment: Motive-group subjects, who know that they will be completing a second task, are, on average, more confident of being in the Performance-Environment than are No-Motive-group subjects who do not know that they have to exert effort in the future.

[2]The signal mimics real life experiences that people may use to infer the importance of effort for success in life: One's colleague may get a promotion even though one considers oneself more talented and productive than the person who got the promotion; other people may get a position to which they applied, knowing that they only got the position because of their personal ties to the company's CEO; still other individuals may win an award for their work knowing that they worked harder and performed better than the other people who were short-listed for the award.

The design, shown in Figure 1, allows me to test for motivating belief distortion non-parametrically by comparing posterior beliefs across the two treatment groups. Nonetheless, I can go further and ask what type of signal leads to motivating belief distortion. First, I can ask whether subjects are more or less likely to engage in motivating belief distortion when receiving a signal that suggests that they are in the Performance-Environment rather than the Chance-Environment. Second, the design allows me to test whether events that are non-informative about the true state of the world affect beliefs. Specifically, I ask whether individuals are more likely to believe in the importance of effort if they won—rather than lost—a reward, holding the informational content of the event constant. This allows me to infer what type of events induce motivating belief distortion.

I further ask whether motivating belief distortion affects decisions about redistribution between two other individuals, as these beliefs are strong predictors of the demand for redistribution for meritocratic individuals. After the first phase of the experiment, in which subjects receive a signal about the environment to which they were assigned (i.e., Chance or Performance) and in which Motive-group subjects are informed about the second task, I give subjects the opportunity to redistribute an initially unequal bonus allocation between two uninvolved participants. These participants were previously recruited to perform the same first task as the decision maker herself. I truthfully tell participants that the initial allocation was determined by the same payment rule they themselves just faced. Subjects can then redistribute this initial allocation. By exploiting variation across treatment groups and signals, I can test (a) whether motivating belief distortion affects inequality-acceptance for other-regarding motives and (b) whether past experiences affect redistributive decisions above and beyond the experiences' informational content.

My results show that subjects distort beliefs to motivate future effort. Subjects who know they will perform another task in the same environment are significantly more confident (seven percentage points) of being in the Performance-Environment. This average effect masks heterogeneity by signal type. Motive-group subjects who received a disincentivizing signal indicating that reward is unrelated to effort, i.e. that they were assigned to the Chance-Environment, are significantly more confident (nine percentage points) that they are in the Performance-Environment compared to control group subjects who received the same signal. My results do not show any difference in beliefs across treatment groups for subjects who received an incentivizing signal indicating that reward is a function of effort, i.e. that they were assigned to the Performance-Environment. This shows that motivating belief distortion is particularly frequent if people receive information that is disincentivizing, i.e., information that indicates that effort is not important for success. By exploiting independent variation in the event that leads to a given signal, I show that motivating belief distortion is particularly pronounced for individuals who know (or believe) that they would have done well in a world that actually rewards effort, i.e., people who learned that the outcome of the task was not justified by their relative performance. Overall, the results of my experiment provide strong evidence that individuals distort their own luck-effort beliefs to motivate themselves for the task they expect to face in the future.

Turning to the results on the distribution decision, I show that motivating belief distortion

does not significantly affect distributive behavior. This suggests that beliefs may be instrumental for motivating future effort but this shift in beliefs is not strong enough to be reflected in aggregate distribution behavior. Even though subjects who are confident of being in the Performance-Environment are less likely to redistribute, I find that past experiences tend to matter a great deal for redistributive decisions in this context. Importantly, I find that subjects who did not win a prize and who performed worse than their competitor redistribute larger amounts than do other subjects, even though the former received a signal indicating a higher likelihood of being in the Performance-Environment. This result highlights the importance of taking into account event characteristics that are not informative about the relative importance of luck and effort when analyzing distributive behavior.

Contribution to the Literature Recent decades have produced a large amount of evidence that beliefs about luck and effort influence attitudes towards redistribution (see Cappelen et al., 2020, for a recent survey). Earlier work, using survey evidence, has demonstrated a robust correlation between the belief that economic inequality is due to luck or effort and the willingness to redistribute (Fong, 2001; Alesina and La Ferrara, 2005; Alesina and Giuliano, 2011). This correlation has been replicated in a more controlled setting using laboratory experiments. These experiments typically vary exogenously regarding whether differences in the initial allocation are due to luck or effort (e.g. Konow, 2000; Cappelen et al., 2007; Krawczyk, 2010; Durante et al., 2014; Lefgren et al., 2016). In these studies, the source of inequality is typically known to the subjects and there is no uncertainty about the true role of luck or effort. Cappelen et al. (2019) look at intensive margin differences in probabilistic luck-effort beliefs by informing subjects about the true probability that success is within one's control. The present paper contributes to this literature by testing empirically how individuals form beliefs about the role of luck or merit and, more specifically, how this belief formation interacts with future incentives faced by the decision maker.

By providing evidence about how individuals form luck-effort beliefs, the paper helps build the micro-foundation of canonical models that explain distributive equilibria. Piketty (1995), for example, studies a model where individuals learn about the relative importance of luck and effort from their own or their ancestors' past mobility experiences. Bénabou and Tirole (2006) study how motivated just-world beliefs may function as a commitment device. Alesina and Angeletos (2005) argue that differences in beliefs originate in historical experiences. Recent work has tested some predictions from these models by exogenously providing individuals with information about mobility and testing the effect of such information on beliefs and preferences for redistribution (Alesina et al., 2018; Gärtner et al., 2019). Gärtner et al. (2019) additionally test and find evidence in support of Bénabou and Tirole (2006)'s prediction that parents transmit distorted beliefs to their children if they expect relatively low levels of redistribution in the future. However, it remains unclear to what extent luck-effort beliefs are used as a motivational device.

Studies using lab experiments have focused mainly on an attribution bias in luck-effort beliefs in a static setting where beliefs have no motivating value (Deffains et al., 2016; Cassar and Klein, 2019; Fehr et al., 2020). These studies found that subjects who lost a contest attribute their

failure to luck and demand more redistribution ex-post, while those who won attribute their success to differences in effort and are less willing to redistribute income. Cassar and Klein (2019) additionally identify an ingroup bias in distributive decisions. A recent experiment by Valero (2020) shows that individuals attribute failure to luck and success to effort but she does not find any evidence that individuals distort beliefs to morally justify the self-serving implementation of low levels of redistribution. Erkal et al. (2021) show that an attribution bias also extends to the evaluation of others' decisions, such that outcomes that are bad for a group are more likely to be attributed to bad decisions while good outcomes are more likely to be attributed to luck. I advance this literature by testing whether luck-effort beliefs are distorted to motivate future effort and whether they are shaped by the incentive structure individuals face. This work, thus, advances our understanding of how the economic environment shapes luck-effort beliefs.

I contribute to a growing literature on motivated belief formation. While the recent empirical literature has made advances in identifying *how* individuals distort beliefs (e.g. Zimmermann, 2020; Chew et al., 2020), we still lack empirical evidence as to *why* individuals want to distort beliefs. Past work has shown that individuals distort beliefs to deceive others (Schwardmann and van der Weele, 2019; Charness et al., 2018) or to justify selfish behavior (Di Tella et al., 2015). Coutts (2019a) and Barron (2020) show that it is unlikely that individuals distort non-ego-relevant beliefs about the likelihood of being in a given state of the world, purely because they expect a higher income in that state of the world. A large part of this literature, starting with Möbius et al. (2014), looks at asymmetric updating after receiving feedback about relative ability and has found mixed results. Convincing explanations for this heterogeneity in results remains lacking (Benjamin, 2019). I advance this literature by explicitly testing whether individuals distort beliefs to motivate future effort while controlling what subjects typically learn about themselves from the feedback they receive. The idea of beliefs as a motivating device was first introduced by Carrillo and Mariotti (2000) and Bénabou and Tirole (2002) to model why individuals are persistently overconfident, which has been shown to have far-reaching consequences, such as excess entry into business (Camerer and Lovallo, 1999).

Few empirical studies explicitly examine whether individuals distort beliefs to motivate effort. Chen and Schildberg-Hörisch (2019) show empirically that higher confidence in one's own ability is related to higher effort provision. This validates a necessary theoretical condition such that there is a demand to distort beliefs about one's own ability for motivating purposes. Banerjee et al. (2020) study the effect of feedback spillovers across unrelated task on confidence in one's own ability to do well in a contest using an artefactual field experiment. While they find evidence that individuals engage in motivated belief distortion, their heterogeneity analysis indicates that this is driven by hedonic rather than instrumental motives. Huck et al. (2018) study performance and information avoidance in the presence of uncertain incentive schemes using a lab experiment. They find preferences for and against information and show that information avoiders outperform information receivers. König et al. (2019) provide field evidence that is consistent with the prediction that beliefs about the return to effort are inflated to motivate effort. Ambuehl (2017) shows that incentives to undertake an unpleasant task in the future affects information

acquisition about how pleasant the task is. The latter three studies are closest to mine because they study how variation in the return to effort affects either information acquisition, beliefs, or behavior. Other studies in this literature, however, hold the motive behind beliefs constant and vary information. My design generates variation in the motive and holds information constant. By varying the motive across individuals, I avoid the problem that the existence of the motive itself is unobserved and endogenous. Importantly, my design allows me to explicitly isolate belief distortion for motivational purposes from affective motives (i.e., deriving utility from the mere fact of believing in one state of the world). Furthermore, I can analyze whether motivating belief distortion interacts with the information that subjects face, as well as with the content of the message, as I vary rank and outcome while holding the informational content of the signal constant.

The remainder of this paper is organized as follows: Section 2 derives the hypotheses I want to test using a simple framework; Section 3 describes the experimental design; Section 4 characterizes the experimental procedures; Section 5 presents the results; and Section 6 concludes.

2 Framework and Hypotheses

To motivate the experimental design, I build a simple framework that mirrors the situation subjects face in the experiment. For an in-depth analysis of a similar setting, see Bénabou and Tirole (2006). The general set-up is borrowed from Bénabou (2015). The aim of this framework is to characterize what generates the treatment effect, what are the main explanatory variables that are necessary to observe motivating belief formation, and what other motives could be relevant in my setting. Note that in this experiment, I do not aim to explain *how* individuals distort beliefs in a motivated way.[3] However, recent contributions to the literature have found strong evidence that individuals forget or recode negative signals about their intelligence (Chew et al., 2020; Zimmermann, 2020).

In this paper, I am interested in testing the hypothesis that, in order to motivate themselves to work hard, individuals overestimate the importance of effort for obtaining a reward. This requires studying a situation where individuals are initially willing to work. A related, but different, hypothesis is that individuals may underestimate the importance of effort in order to morally justify to themselves the decision to not exert effort. This can be rationalized as self-handicapping, i.e., as a strategy to protect one's self-confidence by avoiding the outcome of an activity that is too informative about one's own ability (e.g. Berglas and Jones, 1978). While this latter hypothesis is interesting, it will not be part of the present's work empirical design because it is an affective motive for belief distortion, driven by the demand to maintain a positive self-image, which is not directly related to the demand to distort beliefs for motivating purposes. Furthermore, if this type of belief distortion is relevant in my setting, it would predict the opposite treatment effect. I also will not address purely affective motives for belief distortion

[3] Numerous theoretical models have characterized *how* individuals distort beliefs. In brief, they assume either that agents ignore, forget, or recode past information (e.g. Bénabou and Tirole, 2002, 2006, 2011) or individuals literally choose the beliefs they want to hold and trade them off against the material (Brunnermeier and Parker, 2005) or mental costs (Bracha and Brown, 2012) of belief distortion (see Coutts, 2019b, for an experiment that tests both mechanisms).

(i.e., the desire to believe in a state of the world with higher income prospects). While such motives are certainly interesting to study, they do not capture the notion that individuals distort beliefs to motivate effort, which is the central idea of this paper. In the experimental design section below, I explain how I control for this competing motive.

Agents undertake an activity in a world where the return to their effort is uncertain. Depending on the state of the world, the return θ is either high or low ($\theta_H \geq \theta_L$). I also assume that each state of the world gives the agent an effort-independent payoff κ ($\kappa \in \{\kappa_H, \kappa_L\}$). To relate this setting to the bigger picture, one can say that θ characterizes the degree to which effort is rewarded and κ is the part of income that is independent of effort. κ characterizes, for example, inherited income, genes, or unconditional transfers. Note that for a risk-neutral agent, what matters is how much more she is able to earn from exerting effort. Hence, if her salary Y comprises an unconditional part a and an effort-conditional bonus b, θ would be equal to b. Further, the effect of taxation on the return to effort is already included in θ. I do not model how citizens choose taxation, nor do I explicitly model the underlying redistributive system. I am interested in mapping the return to effort, θ, to belief distortion. These are the parameters that I will manipulate in the experiment.

Agents have time-inconsistent preferences that I characterize using hyperbolic discounting. In period t, they maximize:

$$U_t = \frac{\delta^t}{\beta} U_t + \sum_{\tau=t+1}^{T} \delta^\tau U_\tau \tag{1}$$

δ represents time-consistent discounting while β characterizes the degree of present bias. As β decreases, utility in period t increases. For simplicity, I assume $\delta = 1$ which is what Augenblick and Rabin (2019) actually find for the task that I use.

The framework has three periods: In period 3, the agent receives her payoff $U_3 = \theta e + \kappa$.[4] In period 2, the agent chooses whether or not to exert effort, given her belief p_2. In my experiment, this is a binary decision because I use threshold incentives (subjects earn θ if their effort exceeds a certain threshold), which allows me to control by design for affective motives for belief distortion. Exerting effort e costs the agent $c(e, p)$, where $c(e = 0, p) = 0$. The p term in $c(\cdot)$ characterizes a psychological cost from exerting effort if the agent believes that his effort is wasted or is a "lost cause" and unnecessary. p is the likelihood that her effort actually matters and $c(\cdot)$ decreases convexly in p ($c'_p(e, p) \leq 0, c''_{p,p}(e, p) \geq 0$), meaning that the psychological cost increases with the likelihood that her effort has no effect on the outcome. The cost of effort, both psychological and real, is paid in period 2, while the returns are reaped in period 3. Thus, her decision payoff in period 2 is:

$$\max_e E_2(U(e, c, \theta, \kappa)) = E_2\Big(\theta e + \kappa - \frac{c(e, p)}{\beta}\Big)$$

In the experiment, there are two states of the world, H and L. In period t, the agent believes

[4] In Prolific, where I run the experiment, subjects typically receive the payment after 1-2 days.

that she is in state H with probability p_t and in state L with probability $1 - p_t$. We can then rewrite the problem as

$$\max_e E_2\big(U(e, c, \theta, \kappa)\big) = e\big(p_2\theta_H + (1 - p_2)\theta_L\big) + \big(p_2\kappa_H + (1 - p_2)\kappa_L\big) - \frac{c(e, p_2)}{\beta}. \tag{2}$$

The agent chooses to exert effort in period 2 if:

$$\begin{aligned} E_2\big(U(e = 1, c, \theta, \kappa)\big) &\geq E_2\big(U(e = 0, c, \theta, \kappa)\big) \\ \big(p_2\theta_H + (1 - p_2)\theta_L\big) + \big(p_2\kappa_H + (1 - p_2)\kappa_L\big) - \frac{c(e, p_2)}{\beta} &\geq \big(p_2\kappa_H + (1 - p_2)\kappa_L\big) \\ \big(p_2\theta_H + (1 - p_2)\theta_L\big) &\geq \frac{c(e, p_2)}{\beta} \end{aligned} \tag{3}$$

Note that effort-unconditional income κ cancels out in the decision of a risk-neutral agent. This obviously, would be different if the agent's utility function is concave.[5] The inequality is more likely to hold, the larger p_2 is (because $\theta_H \geq \theta_L$ and because of its psychological effect on $c(\cdot)$) and the larger θ is. It is also more likely to hold if the direct cost of effort c and the degree of present bias β are low. In the experiment, I create exogenous variation in θ_H (across treatment groups). Note that p_2 is the motivated belief and, thus, is endogenous in this model.

We now turn to the behavior of the agent in period 1. The agent in period 1 receives a signal σ that is informative about the true state of the world. She updates her belief in a "non-motivated" way and we denote her posterior p_1.[6] The agent now has the opportunity to ignore, forget, recode, or reinterpret the signal they receive.[7] I assume that this comes at a mental cost $M(p_1, p_2)$. In the spirit of Bracha and Brown (2012), I assume that the mental cost of self-deception is increasing in the absolute *difference* between the non-motivated belief she holds in period 1, p_1, and the belief she recalls in period 2, p_2. Additionally, $M(p_1, p_2)$ can also stand for the cost forgone in the belief elicitation task. Hence, if she maintains her non-motivated belief, she does not incur any mental or material cost from distorting beliefs. A period–1 agent maximizes her expected utility with respect to p_2, which is her choice variable:

$$\begin{aligned} \max_{p_2} E_1\big(U(e, c, \theta, \kappa, p_2)\big) =& e\big(p_1\theta_H + (1 - p_1)\theta_L\big) + \big(p_1\kappa_H + (1 - p_1)\kappa_L\big) \\ & - c(e, p_2) - \frac{M(p_1, p_2)}{\beta} \end{aligned} \tag{4}$$

Note that she maximizes expected utility by anticipating the effect that p_2 has on the period–2 agent's decision about whether to exert effort. The internal solution to (4) is the value for

[5] If utility is concave, the threshold p_2 for which the inequality holds is higher compared to the linear case, *ceteris paribus*. This is due to the fact that the decision maker is now averse to making the mistake of paying the cost of effort without receiving a reward. Furthermore, (3) is less likely to hold for low values of κ_L because losses from exerting effort without receiving a reward now have a large impact on the margin. If κ_H is low and κ_L is high, the opposite is true, as losses have a low effect on the margin, compared to the gains from exerting effort. I want to highlight that the comparative statics that I focus on in this paper remain the same: The inequality in (3) is more likely to hold, the larger p_2 is.

[6] "Non-motivated" means that agents may form non-Bayesian beliefs but they are restricted to not act on the basis of any motive (affective or instrumental).

[7] See the beginning of this section for a discussion concerning the supply side of motivated beliefs.

p_2 that equalizes the marginal gain from belief distortion to its marginal cost $\big(c'_{p_2}(e, p_2) = M'_{p_2}(p_1, p_2), p_2 \in [0, 1]\big)$. Note that this is not necessarily the *optimal* belief, as the optimal belief has to satisfy the constraints an agent faces, which I detail below.

A period-1 agent has two reasons to distort beliefs: first, to counteract a misallocation of effort due to present bias and, second, to ease the psychological burden of working in an environment where it is uncertain whether it is actually worthwhile to undertake this activity. If time-inconsistent preferences are the main driver behind motivating belief distortion, one would only expect the period–1 agent to distort beliefs if she thinks that the period–2 agent is unlikely to work under non-motivated beliefs.[8] If the psychological channel is the relevant mechanism behind motivating belief distortion, then one would still expect subjects to distort beliefs for motivating purposes if they are not sophisticated about their present bias and if the period–2 agent would complete the task under non-motivated beliefs.

We can now ask under which conditions the period–1 agent is willing to distort beliefs. In our simple model, this will be the case if, in period 1, she plans to undertake the task herself in period 2 and if the cost of distorting her beliefs is not too high. Denoting the optimal belief p_2^* (or her initial belief, p_0, if we assume that agents distort by forgetting), we can formulate the necessary condition for distorting beliefs for motivating purposes as:

$$\begin{aligned}
(p_1\theta_H + (1-p_1)\theta_L) + (p_1\kappa_H + (1-p_1)\kappa_L) - c(e=1, p_2^*) - \frac{M(p_1, p_2^*)}{\beta} &\geq (p_1\kappa_H + (1-p_1)\kappa_L) \\
(p_1\theta_H + (1-p_1)\theta_L) - c(e=1, p_2^*) &\geq \frac{M(p_1, p_2^*)}{\beta} \qquad (5) \\
(p_1(\theta_H - \theta_L) + \theta_L) - c(e=1, p_2^*) &\geq \frac{M(p_1, p_2^*)}{\beta}
\end{aligned}$$

The agent distorts beliefs if she would want to work in period 2, evaluated from period–1 agent's point of view (LHS>0), and if the cost associated with distorting beliefs is not too high. Note that the period–1 agent is fully aware that the expected return to effort is a function of p_1 and not of p_2. This property is important for the experimental design, as it implies that motivating belief distortion can only be identified if individuals are, in principle, willing to work under p_1. For example, if $\theta_H = \theta_L = 0$ or if $c(e, p_2 = 1)$ is very high, one should not expect individuals to engage in belief distortion because they would not want to engage in the task in the first place. While this is impossible to test empirically, because we only observe the period–2 agent's decision to exert effort, I can set the incentives for the task such that most of the subjects complete it, which implies that $LHS \geq 0$ in (5). Furthermore, under p_2^*, (3) has to hold as well, which is relevant for agents with low β. For them, the cost of belief distortion may be too high for (3) and (5) to hold simultaneously.

Next, I ask what constitutes p_2^*. As mentioned previously, $p_2^* = p_0$ if subjects distort beliefs by forgetting the signal they received. However, if subjects have more freedom in choosing the beliefs they want to hold, then p_2^* is either the belief that maximizes (4), denoted as p_2^{max}, or it is the minimal p_2 that satisfies (3). The latter is the case if the period–2 agent would not be

[8] In this case, the demand to distort beliefs arises from the need to correct for a misallocation of effort due to present bias. If present bias has no significant impact on the effort chosen, then a subject would not distort beliefs for this reason.

willing to work under $p_2 = p_2^{max}$ but would also be willing under p_2^*, which then has to satisfy both (3) and (5). This would be the case if the agent has a relatively strong present bias (i.e., β is low) compared to the psychological burden characterized by $c(\cdot, p)$.

How does variation in p_1 affect p_2^*? In both cases described above, one would expect more belief distortion if p_1 is low, conditional on (5) holding. If countering present bias is the driving force behind belief distortion, (3) is less likely to hold for low p_1, triggering a higher demand for belief distortion for these subjects. If the psychological cost is the relevant mechanism, we would expect more belief distortion for low p_1 because (3) is less likely to hold under p_1 and because the return to belief distortion is higher under the assumption that $c(\cdot)$ is convex in p.

Variation induced in the experiment This experiment varies p_1 and θ_H, while keeping $\theta_L = 0$. It varies p_1 within treatment groups by varying (Bayesian) posteriors. It varies θ_H across treatment groups: The no-motive-group has $\theta_H = \theta_L = 0$ because it does not know that it will face a task, while for the motive-group $\theta_H > 0, \theta_L = 0$. At first it is useful to note that no-motive-group subjects should never distort beliefs to motivate future effort because (3) never holds (i.e., the return to future effort is always 0 no matter the state of the world because there is no second task). This means that, for no-motive-group subjects, $p_1 = p_2$. Motive-group subjects, however, have an incentive to distort beliefs in order to counter the under-provision of effort due to their time-inconsistent preferences or to weaken the psychological burden of potentially working for no good reason.

I further vary p_1 across subjects. This variation allows me to assess what type of information triggers motivating belief distortion. Because I create a situation where exerting effort is desirable for the period–1 agent, I expect belief distortion to be stronger for subjects with low p_1, i.e. those who received a disincentivizing signal.

This leads me to my first set of hypotheses (H1):

H1.1 Motive-group subjects are, on average, more confident that the state of the world is H than are no-motive-group subjects

H1.2 Belief distortion is more prevalent if the signal is disincentivizing;

H1.3 Subjects with high effort costs or low returns to effort distort luck-effort beliefs.

In the experiment, I test H1.1 by comparing beliefs across treatment groups. I test H1.2 by comparing belief distortion among individuals who received an incentivizing signal with those who received a disincentivizing signal. I test H1.3 by testing whether motivating belief distortion is driven by subjects who either are not very good at the task at hand or do not enjoy the task.

Belief distortion and redistributive behavior The second set of hypotheses tests predictions concerning distributive decisions by spectators with regard to two *other* individuals. To derive these predictions, I build on the evidence that individuals hold normative fairness views and care about the distribution of income across other individuals (see e.g. Almås et al., 2020). I follow the general framework introduced by Cappelen et al. (2013), which was later applied

to a setting very similar to the one studied here (Cappelen et al., 2019), and I refer to their analysis in the remainder of this section.

Concretely, one agent (spectator) is matched with two workers who have different levels of *initial* income. One of the workers (worker H) has an initially higher income y_H^{pre} than the other worker (worker L), i.e. $y_H^{pre} > y_L^{pre}$. The spectator then has the choice to redistribute income from worker H to worker L. The final distribution of income is characterized by the pair $\mathbf{y}^{post} = (y_H^{post}, y_L^{post})$. I further assume that the spectator cares about the final distribution of income and seeks to implement an ex-post distribution of income that is as close as possible to her normative fairness views specified as $\mathbf{y}^{fair} = (y_H^{fair}, y_L^{fair})$. This decision is characterized by the spectator minimizing the following loss function by choosing $\mathbf{y}^{post}$:

$$V(\mathbf{y}^{post}) = -\big(|\mathbf{y}^{post} - \mathbf{y}^{fair}|\big) \tag{6}$$

The literature has identified three common fairness types (see Almås et al., 2020): (1) libertarian fairness ideal, where the spectator considers the initial distribution of income as fair, i.e. $\mathbf{y}^{fair} = \mathbf{y}^{pre}$; (2) egalitarian fairness ideal, where both workers should get the same amount of income, i.e. $\mathbf{y}^{fair} = (\frac{y_H^{pre}+y_L^{pre}}{2}, \frac{y_H^{pre}+y_L^{pre}}{2})$; and (3) the liberal egalitarian or meritocratic fairness ideal, where the worker that produced more units should also get a higher income ex-post, thus $y_H^{fair} > y_L^{fair}$ if worker H produced more units and $y_L^{fair} > y_H^{fair}$ if L produced more units. A meritocrat prefers an equal split if they produced the same amount of units.

In the setting I am interested in, the spectator faces uncertainty about which of the two workers produced more units. While this will not affect the decision made by libertarians or egalitarians, it does affect the decision by meritocrats. Their decision depends on their *beliefs* about which of these two workers was the better performer. These beliefs are affected by spectator's assessment of whether the initial distribution of income reflects differences in effort. If the spectator believes that the initial distribution of income $\mathbf{y}^{pre}$ reflects differences in effort, then $y_H^{fair} > y_L^{fair}$.

If we now assume that there are two states of the world, one where effort matters for the initial allocation of income (true state of the world with probability p) and another where one's income is based entirely on luck (true state of the world with probability $1-p$), then the initial allocation of income serves as a noisy signal regarding who was the better performer. Concretely, the likelihood that the agent with higher initial earnings is also the higher performer increases with p. Following Cappelen et al. (2019), I rewrite the expected value function for meritocratic spectators, where the posterior distribution of income $\mathbf{y}^{post}$ is characterized as the share of total income given to the initially high earner, denoted as s_H, while the total income is normalized at 1. This implies that the share of the initially low earner is $s_L = 1 - s_H$. s_{Best}^{Fair} is the share of income a spectator prefers to allocate to the best performer.

$$\begin{aligned} E\big(V(s_H^{Post}, s_{Best}^{Fair}, p)\big) = &- P(\text{H is best}|p)\big(|s_H^{Post} - s_{Best}^{Fair}|\big)^{\alpha} \\ &- \big(1 - P(\text{H is best}|p)\big)\big(|s_H^{Post} - (1 - s_{Best}^{Fair})|\big)^{\alpha} \end{aligned} \tag{7}$$

The first part in (7) characterizes the loss function for a given amount redistributed if H was indeed the better performer, while the second part characterizes the loss function if L was the better performer. α is an elasticity parameter. Importantly, $P(\text{H is best}|p)$ depends on the true state of the world and the likelihood that the high performer wins the initial prize in either state of the world. One can thus characterize it as

$$P(\text{H is best}|p) = p^M p + p^C(1-p). \tag{8}$$

p^M is the likelihood that a high performer wins the initial prize in the state of the world where effort matters and p^c is the likelihood that a high performaner wins the initial prize in the state of the world where winning is entirely a matter of luck. To capture this notion, I set $p^M > 0.5$ and $p^C = 0.5$ and $P(\text{H is best}|p)$ is then strictly *increasing* in p. This implies that the first part of (7) increases as p increases while the second part decreases, which induces meritocratic spectators to allocate a higher share to H as the likelihood that H is actually the high performer increases. These spectators thus want to further decrease the gap between s_H^{Post} and s_{Best}^{Fair}. The extent to which variation in p leads to inequality acceptance depends on α. For $\alpha = 1$, meritocratic spectators would not redistribute at all, since the initial distribution is informative about who was the better performer. As $\alpha \to \infty$, meritocratic spectators will always redistribute, unless $p = 1$.

In the experiment, I study how a variation in p due to motivating belief distortion affects the redistributive behavior of spectators. I do this by comparing redistributive decisions made by subjects who distort beliefs for motivating purposes and those who do not hold motivating beliefs. Because meritocratically inclined subjects are the only subjects who should react to variation in p, I expect that spectators are, in the aggregate, weakly less willing to redistribute from H to L. Whether the variation in p due to motivating belief distortion is strong enough is an empirical question because it depends on α and also on the share of meritocrats in the population. Furthermore, any potential difference in beliefs across the treatment groups should be mediated through variation in beliefs.

This leads to the second set of hypotheses (H2) that I want to test through my design:

H2.1 Motive-group subjects redistribute less than no-motive-group subjects;

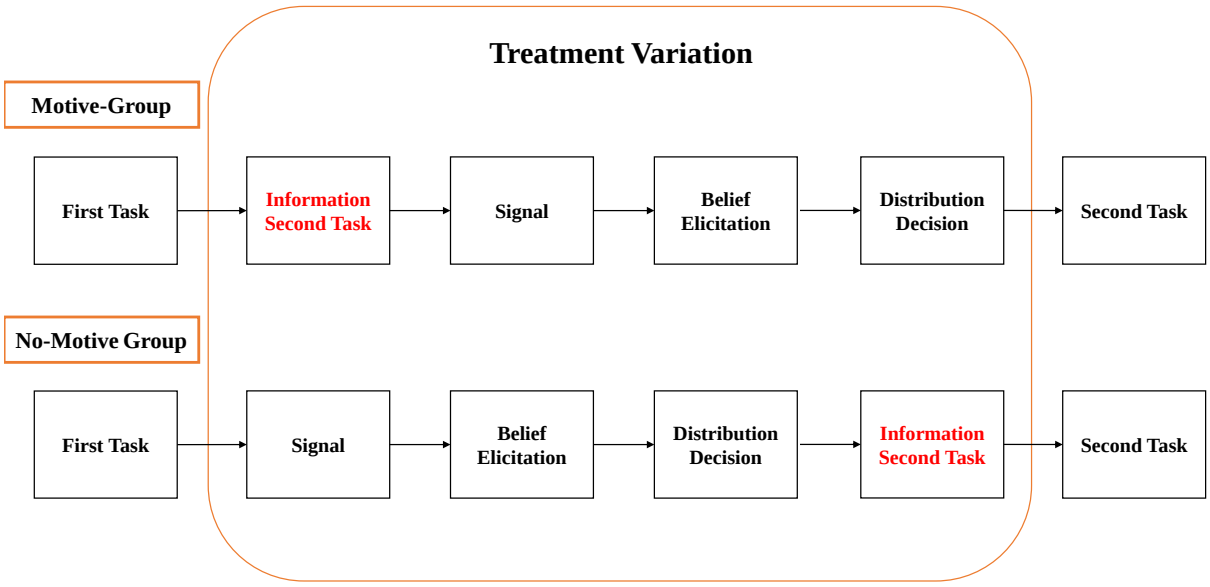
H2.2 The treatment effect on redistribution is mediated by differences in beliefs;

H2.3 The treatment effect is stronger for subjects who hold meritocratic ideals.

As mentioned above, motivating belief distortion may not be strong enough to significantly shift average beliefs around the prior and the relationship through which p affects inequality acceptance may be convex. This means that, on the one hand, evidence in favor of my hypothesis is relatively strong. On the other hand, it means that, in case of a null-result, a more granular design may be needed to identify the relationship between motivating belief distortion and redistributive preferences due to other-regarding reasons.

3 Experimental Design

Figure 1: Schematic diagram characterizing the stages of the experiment



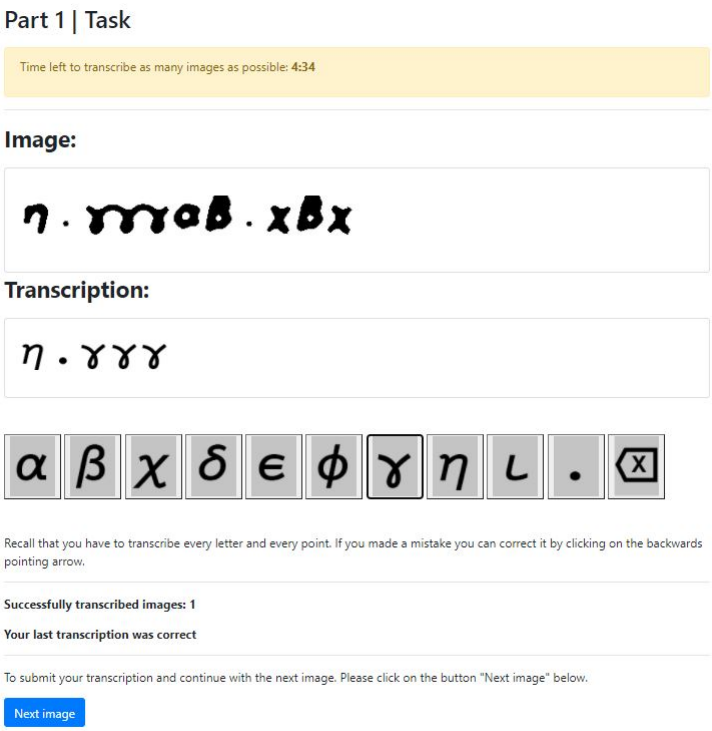
The basic set-up of the experimental design is shown in the schematic diagram in Figure 1. Subjects begin the experiment by engaging in a tedious real effort task (First Task) in an uncertain environment. They either complete the task in a setting where the likelihood of receiving a reward is increasing in their effort (Performance-Environment), or is based entirely on luck (Chance-Environment). After finishing the task, subjects receive a noisy signal informing them whether they are in the Chance- or Performance-Environment. The aim of this signal is to induce variation in baseline beliefs about the state of the world; I then elicit these beliefs. At the end of the experiment, subjects will engage in a second task. As in the first task, the payment rule depends on the environment that was drawn at the beginning of the experiment. The purpose of the second task is to provide a motive to distort beliefs for motivating purposes. Subjects that are informed about the second task may now overestimate the likelihood of being in the Performance-Environment in order to motivate themselves to work hard on the second task. To test for this hypothesis, I vary the point in time at which I inform subjects about the second task. Subjects assigned to the Motive-Group treatment are informed about the second task *before* belief elicitation. Therefore, they have an incentive to distort beliefs for motivating purposes. This is not the case for subjects in my control treatment group, the No-Motive-Group. These subjects are informed about the second task *after* belief elicitation. Therefore, they do not have an incentive to distort beliefs for motivating purposes. I thus vary the motive to distort beliefs for motivating purposes across treatment groups and by comparing beliefs across the two groups, I can identify whether subjects engage in motivating belief distortion. Note that the only difference across the two groups is the point in time at which they are informed about the second task. Importantly, the *objective* information they have about the state of

the world is equally distributed across the groups. To test whether motivating belief distortion affects inequality acceptance due to other-regarding motives, I pair subjects with two agents who previously participated in the first task and, importantly, were in the same environment—Chance or Performance—as the decision maker herself. One of the agents won the bonus in the first task, while the other did not win it. The subject is given the opportunity to redistribute income from the agent who won the initial prize to the other agent. By comparing the amount redistributed, I can infer whether motivating belief distortion affects inequality acceptance due to other-regarding motives. In the following subsections, I present each part of the experiment in detail.

3.1 Work Phase 1

Subjects start the experiment by engaging in a real effort task under (noisy) tournament incentives. The task was chosen because it is tedious and it is relatively independent of skill or intelligence. The incentive scheme was chosen to generate noisy signals about the environment to which the subject was assigned—Chance or Performance—that are not co-linear with effort or with success or failure in the task.

Figure 2: Screenshot from the real-effort task



The task, introduced by Augenblick et al. (2015), consists of transcribing as many images containing 11 Greek letters as possible in 5 minutes. I chose this task for three reasons: First, it is a very tedious task that has been shown to actually generate time-inconsistent behavior (indeed, this is the topic of Augenblick and Rabin (2019), which uses the same task) and may be particularly sensitive to limited willpower. Second, the task does not depend heavily on inherent skills, such as being good in math, which is important because it gives everybody a

realistic chance to complete the task. Third, relative performance on this task may be less likely to affect one's self-image than other tasks used in the literature on overconfidence, such as typical IQ tasks. This third characteristic enables me to generate a benchmark finding in a setting where self-image concerns play a minor role.

Subjects know that they are in one of two environments. If the environment is Performance, there is a positive correlation between winning a prize of €7.50 and exerting more effort. If the environment is Chance, effort is orthogonal to winning the prize. Agents do not know the true state of the world but are informed that they are in either environment with probability 0.5. I chose a prior probability of 0.5 because this prior is easily understood and memorized by subjects and it allows me to generate a sufficiently large share of disincentivizing signals that yield a lower posterior likelihood of being in the Performance-Environment (see below).

Table 1: Likelihood of winning the prize by group rank and environment

	Likelihood to win		
Rank	**Overall**	**If Performance**	**If Chance**
1	0.65	0.8	0.5
2	0.35	0.2	0.5

After completing the task, the subject is matched with one other player (henceforth called "competitor") who completed the identical task in a previous pilot study. As shown in Table 1, the likelihood of winning the prize varies by the environment and rank (whether the subject performed better than her competitor). If the subject is in the Performance-Environment, the likelihood of winning the prize is increasing in the effort she exerted: If she ranks first, the likelihood that she wins is equal to 0.8; if she ranks second, it is equal to 0.2. If she is in the Chance-Environment, the likelihood of winning the prize is always equal to 0.5. Taking both together, the *prior* likelihood of winning if she ranks first is equal to 0.65 and if she ranks second it is equal to 0.35.[9]

3.2 Feedback Phase

After completing the task, subjects receive two pieces of information that compose their signal: (1) Their effort rank within the group and (2) whether they won the prize. *Together*, these two pieces of information enable subjects to form posterior beliefs about the likelihood of being in the Chance- or Performance-Environment.

The structure of the signal is a crucial part of my design because it generates exogenous variation in (non-motivating) baseline beliefs. First, the mechanism will match individuals

[9]It may be that the prior likelihood of winning if ranked first is perceived as too low and the subjects do not want to exert effort in the first place. This type of behavior was, however, quite rare. Only 4.6% of subjects did not transcribe any image correctly in Work Phase 1. Beyond that, subjects are informed how they rank relative to their competitor in the subsequent part of the experiment. Hence, each signal is equally revealing for all subjects.

Table 2: Bayesian posterior likelihood of being in the Performance-Environment

	Outcome	
Rank	**Win**	**Lose**
1	0.62	0.29
2	0.29	0.62

Note. *Rank* is the subject's effort rank within her group. The column *win* (*lose*) is the Bayesian posterior probability of being in the Performance-Environment if the subject won (lost) the prize for every rank.

with equal performance to different competitors. This gives me exogenous variation in being ranked first or second. Second, there is exogenous variation in winning and losing the prize because a random component plays a role in the Performance-Environment. This implies that different subjects may perform equally well at the task but end up with signals that point in opposite directions, allowing me to study the heterogeneity in treatment effects across signals. Furthermore, I observe the signal's *objective* information, which does not depend on unobserved variables such as subjective beliefs about relative performance. This gives me the necessary level of control over baseline beliefs that I need to determine what type of information triggers motivating belief distortion.

Furthermore, both sources of variation enable me to test causally whether individuals update differently after receiving objectively the same information through a message that indicates whether or not they won the bonus. For example, an individual who transcribed fewer images than her competitor may not make the same inference from losing the prize as an individual who transcribed more images than her competitor and wins the prize. This enables me to elucidate whether the message that conveys the signal affects beliefs as such and whether this interacts with my main treatment variation. Seen from a broader perspective, this variation allows me to ask how economic experience, such as failure and success, affects belief in a just world above and beyond its informational content.

In sum, the feedback stage provides subjects with two pieces of information: (1) how many words they transcribed relative to their competitor and (2) whether they won or lost the prize. These two pieces of information together enable subjects to infer the relative likelihood of being in the Chance- or Performance-Environment.

3.3 Information Treatment

Before receiving the signal, half of the subjects are informed about a future part of the experiment. Though the remainder of the experiment is equal across the two treatment groups, their knowledge about it varies: Subjects in the motive-group are informed at this stage that they will engage in a second work-phase where they do the same task as in work-phase 1 but

under threshold incentives (see Section 3.6 for details). Crucially, the environment, drawn at the beginning of the experiment at the subject level, remains constant and matters for work-phase 2: If the subjects are in the Performance-Environment, they win a prize of €5 with probability 1 if they transcribe more than 20 images. If they transcribe fewer images, they do not get the reward. If they are in the Chance-Environment, they win the prize with probability 1, independent of the number of letters they transcribe. Hence, their expected earnings are equal across both states of the world if they plan on completing the task. This alleviates concerns that subjects distort beliefs because they anticipate a higher income in one environment compared to the other. Subjects in the no-motive-group face the identical task but are informed about this *after* the belief-elicitation stage.

This treatment variation enables me to identify whether *elicited* beliefs are distorted by the motive to motivate future effort because it provides a "clean counterfactual"(see Coutts, 2019a, for a methodological discussion). The experiment is identical for both groups up to the point that motive-group subjects are informed about the next task. Importantly, the objective information that subjects receive is equally distributed across the two treatment groups. Hence, even if subjects engage in other forms of non-Bayesian belief updating, a difference between the two treatment groups can only be explained through the fact that motive-group subjects know that they will have to redo the task in the same environment as before. This makes my results robust to functional form assumptions. Most importantly, priors and signals are balanced across treatment groups.[10]

While my hypothesis is that the treatment increases the motivational value of beliefs, it may also be the case that my treatment yields excuse-driven behavior that leads to a *negative* treatment effect (rather than the hypothesized positive effect). As briefly mentioned at the beginning of Section 2, subjects may want to bias their belief downwards in order to morally justify the decision not to work. This should be relevant for motive-group subjects who do not exert any effort at all or maintain positive self-confidence. This is, however, rarely the case, as I show at the end of Section 3.6.

Note that the motive-group subjects receive the information *before* they receive the signal. This implies that I equalize the time between learning about the task's outcome and belief elicitation across subjects. This alleviates concerns that a potential treatment effect can be rationalized by differences in the time that has passed between the reception of the signal and belief elicitation.

3.4 Belief Elicitation Phase

In the next part of the experiment, I elicit subjects' probabilistic beliefs about being in the Performance-Environment. Belief elicitation takes place before the redistribution stage (see above). Hence, at the point of belief elicitation, motive-group subjects are already informed about the second task, but this is not the case for no-motive-group subjects. To elicit beliefs, I

[10] My information condition may induce motive-group subjects to take on an external gaze and make more rational assessments of beliefs because they are primed to think more about the true state of the world. While this could indeed be a difference across the treatment groups, it would imply that motive-group subjects are more reactive to the treatment, which goes against my hypotheses.

follow Schwardmann and van der Weele (2019), who use a variant of the BDM to elicit beliefs. Importantly, the method is incentive-compatible to risk-aversion and asks subjects to choose a likelihood between zero and one-hundred that is divisible by ten.[11] Following Danz et al. (2020), I chose to not instruct subjects about the details of the payment mechanism, but I simply instructed them that it is in their best interest to select the likelihood that is closest to what they truly believe to be true. They can, however, click on a button that gives them detailed information about the payment mechanism and 23% of subjects did eventually click on the button. Subjects can earn €2.5 for this decision.

3.5 Redistribution Stage

To test how motivating belief distortion affects redistributive behavior among two *other* individuals, I ask subjects to redistribute income between two workers who were previously recruited on Prolific to do the same transcription task under the same payment rule as they faced themselves in work-phase 1. Building on the design used in canonical spectator experiments (e.g. Almås et al., 2020), I inform the subjects that one of the two workers with whom they are matched received a bonus of €4, while the other did not receive the bonus.[12] The two workers were each others' competitors and this information was explicitly communicated to the subjects. This implies that the subjects know that if they are in the Performance-Environment, the worker endowed with the initial bonus transcribed more images with 80% probability; if they are in the Chance-Environment, the worker endowed with the initial bonus transcribed more images with 50% probability. Beyond that, subjects are informed that the two workers do not know who earned the original bonus, nor did they know the state of the environment in which they performed the task. Subjects are also instructed that their decision will remain anonymous.

After reading the instructions, subjects could redistribute the initial endowment among the two subjects. Subjects had the option to either not redistribute at all, redistribute €1 from Worker A to Worker B, redistribute €2 from Worker A to Worker B, redistribute €3 from Worker A to Worker B, or redistribute €4 from Worker A to Worker B. After a subject made her decision, I ask to what extent she agrees with the statement that the worker that transcribed more images should also be paid more. This gives me a broad measure of whether the subject is a meritocrat.[13]

Observing subjects' distribution decision allows me to ask whether motivating belief distortion affects preferences for redistribution due to other-regarding motives. This can be done non-parametrically by comparing average amounts redistributed across the two groups. The

[11] Subjects are asked to indicate a probability P that is divisible by ten and makes them indifferent between winning a monetary prize with probability P and receiving the same prize if they are in the Performance-Environment. After having indicated this probability, the computer randomly chooses an integer x that is divisible by ten and lies between zero and one-hundred. If x is higher than P, the subject will participate in a lottery where she wins the prize of €2.5 with probability x. If x is lower than P, she wins the prize if she actually is in the Performance-Environment.

[12] The bonus is smaller than the one the subjects could earn themselves because the experiment involving the workers was shorter in duration.

[13] In the pilot study, I asked subjects how they would redistibute the endowment knowing that they were in the Performance- or the Chance-Environment using the strategy method. In the main experiment, I refrained from this to avoid diluting the incentives for the redistribution decision of interest.

comparison across the two groups should capture the effect of motivating belief distortion on inequality acceptance due to other-regarding motives.[14] I can also go further and conduct exploratory analyses where I ask whether the type of feedback affects distributive decisions. Importantly, I can analyze whether experiencing success oneself affects distributive decisions among others.

3.6 Work-Phase 2

Subjects engage in a second work phase after belief elicitation. This task serves as an incentive to distort beliefs for motivating purposes. The idea is that subjects distort baseline beliefs to motivate themselves to complete this task. The real-effort task is identical to the one used in Work-Phase 1 but incentives are different. The incentives are designed with three goals in mind that allow me to isolate the motive to distort beliefs for motivating purposes: First, the payment rule depends on the environment drawn at the beginning of the experiment; second, the incentives have to be strong enough such that all subjects want to engage in the second work phase to allow for the existence of the motive; and third, I want to equalize expected payoffs across the two environments, conditional on having completed the second task, to exclude the alternative hypothesis that subjects may prefer to believe that they are in the environment with higher income prospects.[15]

Thus, I chose to use the following incentive scheme for the second work-phase: Subjects have up to ten minutes to transcribe twenty images. The payment scheme depends on the environment drawn at the beginning of the experiment and that remains constant throughout the whole experiment. Subjects win a bonus of €5 with probability 1 if they are in the Performance-Environment and completed the task. If they do not complete the task, they will not win the bonus. They win the bonus with probability 1 if they are in the Chance-Environment, no matter how many words they transcribed. If they do not want to complete the task, they have the opportunity to quit the task and proceed with the post-experimental questionnaire.

I use threshold incentives because it allows me to control expected gains from exerting effort in a very salient manner. Importantly, the payoff is equalized across the two environments, conditional on planning to undertake the task. Furthermore, threshold incentives make the effort decision binary. Hence, I know whether the participation constraint, defined in my framework, holds or not. This is essential because it allows me observe whether subjects were actually willing to work on the second task, even if they received a disincentivizing signal.

Due to this design choice, I am not able to study the effect of motivating belief distortion on effort choices. As mentioned above, I can only identify motivating belief distortion for subjects who initially want to complete the second task, even if they received a disincentivizing signal.[16]

[14] One should take into account that the two groups differ in that motive-group subjects know that they will engage in another task and this may affect their preferences for redistribution. Specifically, motive-group subjects face higher income prospects for the rest of the experiment, which could have a countervailing effect on inequality acceptance. This would be a particular concern if most subjects decide to fully redistributed in the experiment.

[15] I want to control for this affective motive for belief distortion, which is different from the kind of motivated belief distortion that is the focus of this study.

[16] To put it differently, the period-1 agent in my framework, who holds non-distorted beliefs, should always want to attempt to transcribe the twenty images in Work-Phase 2.

This implies that I do not expect a difference in effort across the two groups and, for that reason, I did not include an analysis of Work-Phase 2 in my pre-analysis plan.

The notion that exerting effort in Work-Phase 2 is desirable for all subjects is further supported by my data. 93% of subjects did not exit the task prematurely. This shows that subjects are willing to attempt to reach the threshold. This is further supported by looking at the attempts subjects made. More than 90% of subjects attempted to transcribe at least 20 images and more than 95% attempted to transcribe at least 5 images.

3.7 Post-experimental questionnaire

At the beginning of the post-experimental questionnaire, I ask subjects to recall (a) their rank and (b) whether they won the bonus for Work-Phase 1 or not. Subjects also had the option to state "I do not recall." This line of questioning allows me to observe whether subjects are more prone to forget some signals than others. This information also sheds light on the supply-side mechanism of motivating beliefs.

To obtain an incentivized measure of whether subjects use commitment devices to motivate future effort, I offer subjects the opportunity to perform the same task two weeks later under a piece-rate incentive scheme. They can earn up to €10 by transcribing 40 images. Crucially, if they transcribe fewer strings, they receive a partial payment because each string is rewarded with €0.25; i.e. if they transcribe 10 images, they receive €2.50. I first ask subjects how many images they would wish to transcribe if they were reinvited. I then give them the chance to choose a minimum effort level (minimum number of transcribed letters). If they transcribe fewer images than this threshold, they earn nothing. If they transcribe more images, the usual piece-rate scheme is applied. Hence, subjects can make it costly for their future selves if they choose to work less than the amount specified in their minimum production level. Putting both (plan and commitment) together allows me to construct a measure of demand for commitment.[17] The decision is implemented for 5% of subjects.

Subjects are then asked to fill out a post-experimental questionnaire. The questions cover demographics, risk and time preferences, and two attention checks. The question items are provided in Appendix C.

3.8 Summary of the variation generated through the design

Before going through the results of the experiment, it is worth taking a step back to summarize the exogenous variation generated through my design. My main treatment variation is whether subjects have an incentive to distort beliefs for motivating purposes at the point of belief elicitation. The treatment is randomized at the subject level and each subject had an equal probability of being in the motive- or no-motive-group. In the experiment, 248 subjects were assigned to the no-motive-group, while 252 were assigned to the motive group. This variation is used to identify a motive-group effect.

[17] A contract in the same spirit was studied in the field by Kaur et al. (2015). They show that there is a demand for this type of "dominated" contract and that it significantly increases productivity and earnings.

Across subjects, I further vary the feedback that subjects receive. There, I exploit two sources of exogenous variation: First, winning or not winning the bonus is, by design, random once I take into account the performance rank. This means that two subjects may both be better or worse than their competitor but one ends up winning the bonus while the other does not win the bonus. Second, being ranked first or second is exogenous, once I control for the subjects' performance. The intuition is that two equally talented subjects may both win or lose the bonus but end up being ranked first and second because they got matched to different competitors.

This yields variation within treatment groups to obtain an incentivizing and a disincentivizing signal. Overall, 35% of subjects receive an incentivizing signal, while 65% of subjects receive a disincentivizing signal.[18] Holding the informational content of the feedback constant, it is equally likely that a subject receives an incentivizing (disincentivizing) signal through a message stating that she won the bonus and ranks first (second) or that she lost the bonus and ranks second (first). I exploit this variation in my analysis to test whether the informativeness of the signal interacts with my treatment effect.

4 Procedure

Subjects proceed through every part of the experiment (Work-Phase 1, Belief Elicitation, Redistribution, Work-Phase 2). At the end of the experiment, one part will be randomly drawn to become payoff relevant.[19] Subjects know that there are multiple parts at the beginning of the experiment and each part is explicitly introduced and concluded.

All subjects were recruited using Prolific (see Palan and Schitter, 2018, for information about using Prolific as a subject pool). Subjects could only participate in the experiment once. The interface was programmed using oTree (Chen et al., 2016). I only invited subjects who reside in the US or Great Britain and are citizens of either country. This ensured that subjects were fluent in English and capable of understanding the instructions. Beyond that, I only invited subjects who have completed 10 or more surveys and whose submissions had been accepted at least 95 percent of the time. As pre-specified, I do not include subjects in my analysis who did not pass the two attention checks. Furthermore, all subjects who finished the study passed the first comprehension test in no more than two tries. Subjects were informed before they started the comprehension test that they cannot continue with the study if they do not complete the task in at most two trials.[20]

In total, 500 subjects completed the experiment and passed both comprehension checks. They received a €3 show-up fee for completing the study. The median time to complete the study

[18] Motive-group subjects are weakly more likely to receive an incentivizing signal ($p = 0.09$, Fisher's exact test). Note, that this does *not* explain the difference across treatment groups as shown in Section 5.3, where I discuss heterogeneity.

[19] One interesting implication of my findings is that subjects may overestimate the likelihood that a high-stakes part of the experiment is actually implemented.

[20] Subjects who did not pass the comprehension test were paid a €0.50 show-up fee to reimburse them for the (short) time they spent with the initial instructions. To avoid selection effects, they were informed about this *after* failing the comprehension test twice. Overall, only 37 subjects out of 550 who attempted the comprehension test failed to answer all questions correctly after the second try.

was 30 minutes (as announced in the initial description of the study). The median earning in the experiment was €5.50. The experiment was implemented between September 18 and September 20, 2020. Twenty subjects were reinvited to perform the same task under a piece-rate scheme on October 4.

Table B2 in the Appendix presents subjects' individual characteristics. Subjects were on average 33 years old. 40% of subjects are female, 24% are students, and 43% work full time. Table B2 presents a balancing table across the motive- and the no-motive-group. Covariates are, overall, well-balanced across the treatment groups. Motive-group subjects are, however, slightly more risk seeking ($p = 0.062$, Wilcoxon rank-sum test). Furthermore, Table B3 in the Appendix presents results for balancing tests by incentivizing and disincentivizing signal. All covariates are balanced on this dimension. Table B4 in the Appendix presents results for balancing tests by winning or not winning the bonus. Respondents who won are slightly older ($p = 0.1002$, Wilcoxon rank-sum test; $p = 0.083$, t-test). This difference becomes insignificant, however, once I control for performance on task 1 ($p = 0.739$). Furthermore, female subjects were more likely to win the bonus ($p = 0.068$, Fisher's exact test; $p = 0.056$, t-test). This correlation is likely to be spurious as the difference is nearly the same after conditioning on task 1 performance ($p = 0.068$). That the latter correlation is likely spurious is further supported when comparing rank by gender as shown in Table B5 in the Appendix. Female subjects are not more likely to be ranked first ($p = 0.628$, Fisher's exact test; $p = 0.648$, t-test) and this correlation further weakens after conditioning on task 1 performance ($p = 0.885$, Fisher's exact test). Subjects who rank first are, however, significantly younger ($p < 0.001$, Wilcoxon rank-sum test). This correlation remains significant at the 10 percent level if I control for task 1 performance and is likely to be driven by a subset of younger subjects who outperformed the competitor sample (and hence were always ranked first). The other covariates are all balanced across ranks. Throughout the text, I will refer to regressions that control for observed heterogeneity, including risk aversion, age, and gender. I refer to the corresponding tables that are found in Appendix B throughout the analysis. I additionally replicate tables that answer the central research questions for a sample that excludes people who did not answer a simple question about the instructions at the *end* of the experiment.[21] I refer to these tables in the table notes.

Finally, I additionally recruited 250 subjects to serve as workers for the redistribution task. These workers were recruited *before* the main experiment was implemented. Workers were paid a €1.00 show-up fee. They were only hired to do the 5-minute transcription task (Work-Phase 1) and were instructed that the initial distribution of the bonus might be redistributed by a third party. Workers received the implemented bonus after the completion of all sessions.

[21] The question asks: *The likelihood of receiving the bonus in Part 1 of the survey was independent of your performance if you were in the (a) Performance-Environment, (b) Chance-Enviornment, (c) None of the above.* 81 percent answered correct. The share of subjects that answered correct is not significantly different across treatment groups ($p = 0.409$, Fisher's exact test).

5 Results

In the following analysis, a signal is defined as incentivizing or disincentivizing purely based on its *informational* content. Hence, I say that a subject received an incentivizing signal if she learned that she ranked first and won the bonus or if she ranked second and did not win the bonus. Contrarily, I say that a subject received a disincentivizing signal if she won the bonus but was ranked second or if she did not win the bonus though she was ranked first.

The experimental design and analysis was pre-registered. The pre-analyis plan is provided in Appendix E. Throughout the text, I will address whether and how I diverge from the pre-registered analysis. One general divergence is the use of non-parametric Wilcoxon rank-sum tests instead of t-tests, as the former perform better in smaller samples.

5.1 Are signals informative?

Before analyzing differences between the motive- and the no-motive-group, it is worth stepping back and asking to what extent baseline beliefs reflect information contained in the signals.

Figure 3: Mean elicited beliefs by signals (no-motive-group subjects only)

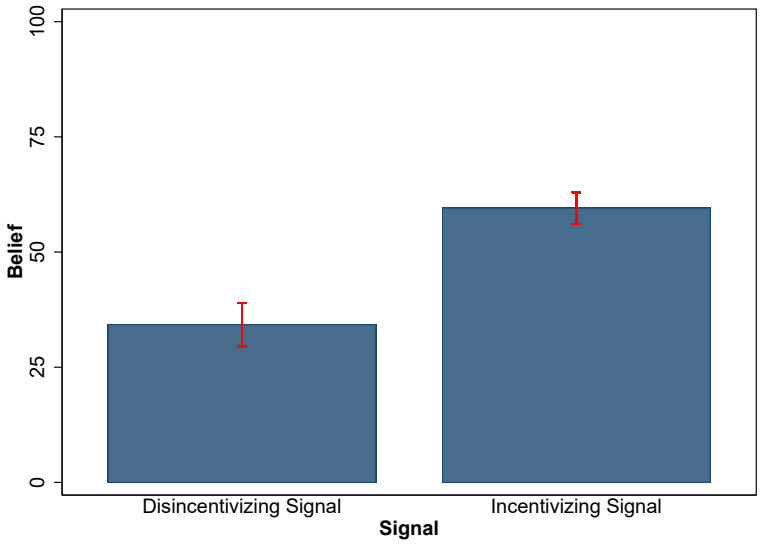
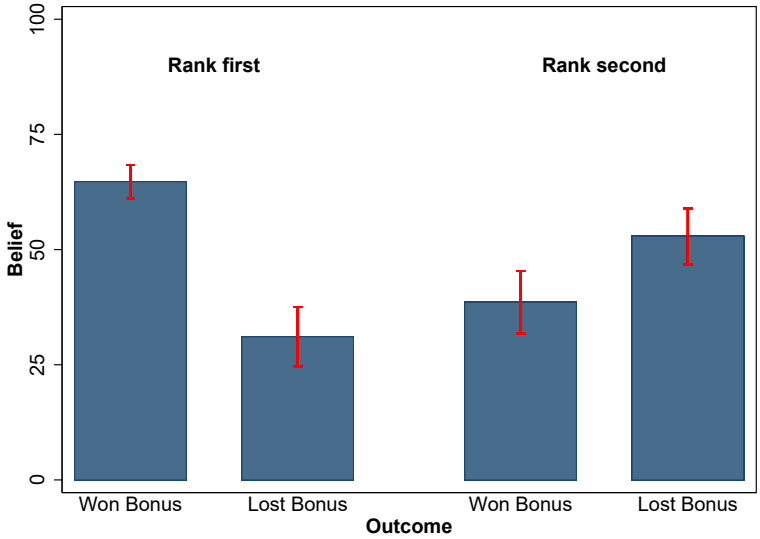


Figure 3 shows average beliefs by signal type for no-motive-group subjects. No-motive-group subjects who received an incentivizing signal are 25 percentage points more likely to believe that they are in the Performance-Environment than are the no-motive-group subjects who received a disincentivizing signal ($p < 0.001$, Wilcoxon rank-sum test). No-motive-group subjects who received a disincentivizing signal believe, on average, that they are in the Performance-Environment with 34 percent probability and those who received an incentivizing signal believe themselves to be in the Performance-Environment with 59 percent probability, on average. Both groups are, on average, slightly more conservative than the Bayesian prediction (0.62 and 0.29). Though the distribution of beliefs is relatively noisy, the difference is supported by looking at the full distribution of beliefs. Beliefs of subjects who received a disincentivizing signal are shifted

to the left of 50%, while those of the subjects who received an incentivizing signal are shifted to the right (see Figure A1 in the Appendix).

Figure 4: Mean beliefs by winning and losing the bonus and rank (no-motive-group subjects)



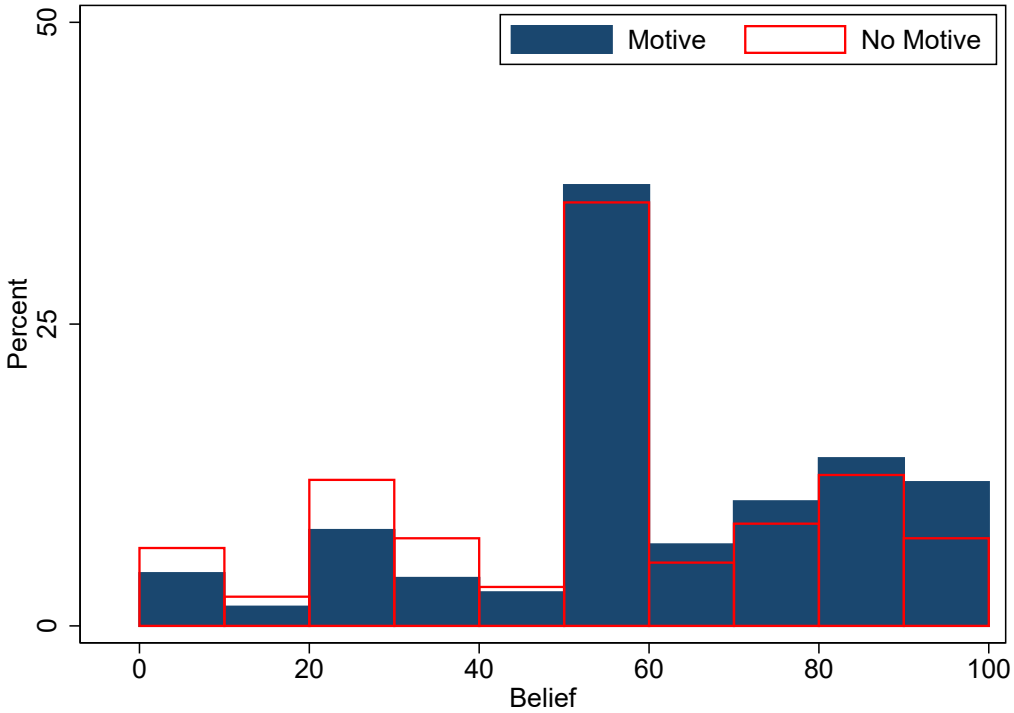
This difference in beliefs remains significant and meaningful when comparing beliefs across subjects who received a different signal and stratifying by whether the subjects ranked first or second, as shown in Figure 4. Subjects who did not win the bonus and ranked second are 20 percentage points more optimistic about being in the Performance-Environment than are subjects who did not win the bonus but ranked first ($p < 0.001$, Wilcoxon rank-sum test). Similarly, I find that subjects who won the bonus and ranked first believe with 26 percentage points greater probability that they are in the Performance-Environment than do subjects who won the bonus but ranked second ($p < 0.001$, Wilcoxon rank-sum test). Hence, the information subjects extract from the signal is similar in both cases, but the difference seems starker if the subject ranks first in terms of performance. I will return to the effect of winning the bonus later in the paper when I discuss how it interacts with my main treatment variation. Overall, the evidence shows that subjects do, on average, take information provided by the signal into account. Table B6 replicates the results presented in this section controlling for observed heterogeneity.

5.2 Beliefs across motive- and no-motive-group

Figure 5 shows the distribution of beliefs by treatment group for the whole sample. The blue bars show the distribution of motive-group subjects' beliefs while the transparent bars with the red frame show the distribution of no-motive-group subjects' beliefs.

The plot shows that motive-group subjects believe, on average, that it is more likely that they are in the Performance-Environment than in the Chance-Environment compared to no-motive-group subjects. The distribution is slightly shifted to the right and this shift is reflected in average beliefs. Motive-group subjects believe, on average, that they have a 56% probability of being in the Performance-Environment, while for no-motive-group subjects, on average, the

Figure 5: Distribution of beliefs by motive- and no-motive-group subjects (full sample)



probability is 49%. Hence, motive-group subjects are 7 percentage points more optimistic about being in the Performance-Environment than are no-motive-group subjects ($p = 0.004$, Wilcoxon rank-sum test).[22]

As shown in my framework, I expect motivating belief distortion to be stronger if the subject receives disincentivizing rather than incentivizing information. To test for this hypothesis, I compare beliefs across treatment groups, stratifying by the information subjects received.

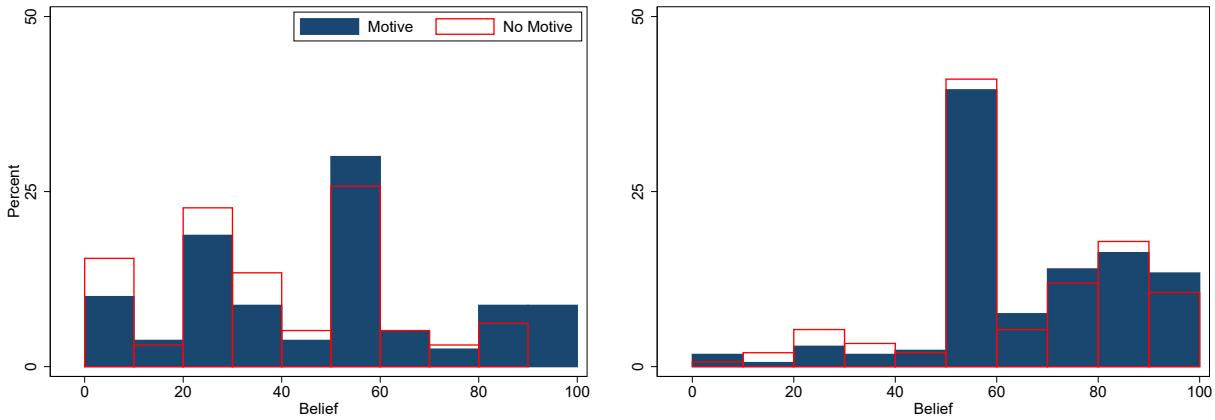
Figure 6 plots the distribution of beliefs across treatment groups by subjects who received either a disincentivizing—sub-figure 6(a)—or an incentivizing signal—sub-figure 6(b). We start by comparing beliefs across subjects who observed an incentivizing signal. Sub-figure (b) shows that the differences across these groups is rather small and there is no systematic shift in the distribution. Overall, the difference in mean beliefs among these subjects is small. Motive-group subjects are 3 percentage points more optimistic about being in the Performance-Environment and I cannot reject the null-hypothesis that beliefs are equal across the two groups ($p = 0.28$, Wilcoxon rank-sum test).[23] Overall, this suggests that there is no systematic difference in updating behavior across treatment groups among subjects who received an incentivizing signal.

[22]Interestingly, average beliefs of no-motive-group subjects are not significantly different from 50% ($p = 0.8220$, t-test, c.i. = [46.5, 52.8]). Hence, there is no evidence that subjects have a general tendency to believe that they have control over their outcome because they derive comfort in believing that they have control over their own income, as postulated by theories that argue for affective motives in a just world (e.g. Lerner, 1980).

[23]This is further supported by comparing the share of subjects who (falsely) updated negatively after receiving an incentivizing signal. Even though the share of subjects who updated negatively is slightly higher (4 percentage points), I cannot reject the null-hypothesis that motive-group subjects are less likely to update negatively than no-motive-group subjects ($p = 0.172$, Fisher's exact test, one-sided) among subjects who received an incentivizing signal.

Figure 6: Distribution of beliefs by motive- and no-motive-group subjects across signals

(a) Disincentivizing signal (b) Incentivizing signal



Note. These figures plot the distribution of beliefs by treatment group. Sub-figure (a) plots the distribution for subjects who received a disincentivizing signal. Sub-figure (b) plots the distribution for subjects who received an incentivizing signal.

Sub-figure 6 (a) plots the distribution of beliefs by treatment groups among subjects who received a disincentivizing signal. The shift in the distribution indicates that motive-group subjects were less likely to update their beliefs negatively compared to no-motive-group subjects who received a disincentivizing signal. This is supported when comparing the share of subjects who updated negatively across treatment groups; 59 percent of no-motive-group subjects updated negatively, while only 45 percent of motive-group subjects updated negatively. Hence, motive-group-subjects are 14 percentage points less likely to update in the correct direction. This difference allows us to reject the null-hypothesis that motive-group subjects are more likely to update negatively than are no-motive-group subjects ($p = 0.035$, Fisher's exact test, one-sided). This dynamic is also reflected when comparing average beliefs of subjects who received a disincentivizing signal across treatment groups. Among this group of subjects, no-motive-group subjects, on average, believed themselves to be in the Performance-Environment with 34% probability, while motive-group subjects, on average, believed this with 43% probability, leading to a significant 9-percentage-point difference in beliefs across the two groups ($p = 0.008$, Wilcoxon rank-sum test).

In sum, my data provide strong evidence that motive-group subjects are significantly more likely to believe that they are in the Performance-Environment than are no-motive-group subjects. Motive-group subjects, on average, believe with higher probability that they are in the Performance-Environment than do no-motive-group subjects and the former's average belief is significantly higher than 50% ($p < 0.001$, t-test). This difference is especially pronounced for subjects who received a disincentivizing signal. Tables B7 and B8 in the Appendix confirm the main results of this section using linear regression and controlling for individual heterogeneity.[24] Given that, by design, 65% of subjects received an incentivizing signal, it is remarkable to find a

[24]The motive-group treatment effect for subjects who received a disincentivizing signal becomes marginally significant ($p = 0.074$, two-sided) once I control for observable heterogeneity. Note, this is driven by limited power when regressing the treatment separately for subjects who received a disincentivizing and incentivizing signal. As

significant average effect for all subjects. Overall, motivating belief distortion may be even more important for average beliefs, if individuals hold false beliefs about their relative ability. Indeed, overconfident beliefs about one's relative ability will generate disincentivizing signals about the true importance of effort if subjects refuse to update about their own relative ability.

Updating relative to a Bayesian benchmark Comparing the belief distortion across subjects who received an incentivizing or disincentivizing signal is complicated because disincentivizing signals are more informative about the true state of the world than are incentivizing signals. I circumvent this issue by characterizing motivating belief distortion relative to a Bayesian benchmark. To that end, I follow the experimental literature on motivated belief updating by calibrating a quasi-Bayesian updating regression (see e.g. Möbius et al., 2014; Coutts, 2019a; Barron, 2020). The outcome variable of the regression is the elicited logit belief ($\tilde{\pi}_i$) at the subject level and the explanatory variable is the log-odds ratio of the signal a subject receives.[25] If a subject is Bayesian, the coefficient of the log-odds ratio would be equal to 1, indicating that her posterior does not differ from the posterior of a Bayesian updater. If the coefficient is larger than 1, she is updating more strongly in response to the signal than she would if she were a Bayesian updater. The latter is an example of a subject who over-responds to the signal. If the coefficient is smaller than 1, she is conservative and does not take all the information from the signal into account when updating beliefs. The coefficient of the prior indicates whether the subject's belief updating is characterized by base-rate neglect or confirmatory bias. My design cannot inform about the latter biases in belief updating because prior beliefs are equal to 0.5 for all subjects. Hence, logit prior beliefs are equal to 0 and the coefficient cannot be identified.[26]

To characterize updating behavior relative to the Bayesian benchmark across incentivizing and disincentivizing signals (q), as well as across motive-treatment groups, I interact the log-odds ratio ($\tilde{q}$) with a dummy variable that indicates whether the subject received an incentivizing or disincentivizing signal. This results in two explanatory variables that characterize updating behavior relative to a Bayesian benchmark for subjects who received an incentivizing or disincentivizing signal separately (indicated by the subscript + and −). Furthermore, I add an additional interaction term, where I interact $\tilde{q}_{+/-}$ with a motive-group (M) indicator. The coefficient of the resulting variable informs whether motive-group subjects are more or less likely to respond to the signal than are no-motive-group subjects. (9) characterizes the resulting model that can be estimated using OLS:

$$\tilde{\pi}_i = \alpha_+ \tilde{q}_+ + \beta_+ \tilde{q}_+ M + \alpha_- \tilde{q}_- + \beta_- \tilde{q}_- M \tag{9}$$

In total, the regression yields four parameters: α_+ and α_- inform how no-motive-group subjects

shown in Table B7, the motive-group effect remains unchanged (both in magnitude and significance) if I control for observable heterogeneity using the complete sample.

[25] Note that logit-beliefs converge to (minus) infinity for subjects who believe that they are in the Performance-Environment with probability 1 (0). To include these subjects in my regressions, I assume that they hold probabilistic beliefs that are equal to .99 (.01).

[26] Note that the aim of the paper is to study differences in belief updating due to the motive treatment rather than to provide a comprehensive characterization of belief updating. To that end, one should elicit beliefs after multiple signals.

update relative to a Bayesian benchmark if they receive an incentivizing and disincentivizing signal. β_+ and β_- inform about how motive-group subjects respond to an incentivizing (disincentivizing) signal compared to the response of no-motive-group subjects who received the same signal.

Table 3: Results of the updating regression specified in (9)

Coefficient	**Estimate**
α_+	1.197
	(0.244)
β_+	0.424
	(0.361)
α_-	1.176
	(0.193)
β_-	-0.787**
	(0.322)
$H_0 : \alpha_+ = \alpha_- = 1$	$F(2, 496) = 0.74$
$H_0 : \beta_+ = \beta_-$	$F(1, 496) = 6.27$
N	500
R^2	0.16

Note. This table shows coefficient estimates for the regression characterized in (9). α coefficients test the null-hypothesis that $\alpha = 1$ while β coefficients test the null-hypothesis that $\beta = 0$. Table B21 in the Appendix replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Robust standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table 3 presents the results of this regression. The log-odds ratios indicate that no-motive-group subjects do not significantly depart from Bayesian posteriors. The null-hypothesis that the coefficient is equal to 1 cannot be rejected for no-motive-group subjects who received an incentivizing ($p = 0.420$, F-test) or disincentivizing signal ($p = 0.361$, F-test).[27]

The interaction terms with the treatment dummy variable show that motive-group subjects who receive an incentivizing signal react more strongly to the signal than do no-motive-group subjects who received an incentivizing signal. The coefficient is, however, not significantly different from zero ($p = 0.240$). This changes when characterizing updating behavior by subjects who received a disincentivizing signal. They react less strongly ($p = 0.015$) to the signal than do no-motive-group subjects who received the same signal. Combining the main effect with the interaction term results in a posterior logit belief of -0.38, which is significantly smaller than the posterior they would hold based on Bayesian beliefs ($p = 0.018$, F-test) and it is not significantly

[27]This does not mean that no-motive-group subjects update similarly to Bayesian agents on the individual level, but the results reflect average updating behavior.

different from 0 ($p = 0.132$, F-test). The latter means that I cannot reject the null-hypothesis that subjects do not update at all.

The next step is to compare the two interaction terms with each other. I reject the null-hypothesis that the two terms are equal ($p = 0.013$, F-test). This reveals significant differences in how treated subjects react to information after receiving an incentivizing or disincentivizing signal. The negative sign of β_- indicates that motive-group subjects *under*-react to the information they received if they received a disincentivizing signal. This is not the case for motive-group subjects who received an incentivizing signal. There, the positive, but insignificant, β_+ indicates that subjects are not updating differently than the baseline; if anything, these subjects are over-inferring from the information. This dynamic is predicted by models of motivated belief distortion: i.e., subjects who receive information that goes *against* the motivated belief under-react to the information they receive, while those who receive information congruent the motivated belief are expected to weakly over-react to this information Bénabou (2015).[28]

5.3 Who distorts luck-effort beliefs to motivate effort?

The previous section established that subjects do distort beliefs to motivate future effort and that motivating belief distortion is particularly pronounced if subjects receive a disincentivizing rather than an incentivizing signal. First, I analyze whether motivating belief distortion interacts with the event observed by the subjects and then I characterize what type of subjects is more prone to distort luck-effort beliefs to motivate effort.[29]

The effect of winning the bonus and performance rank on beliefs To analyze how winning and losing affects beliefs, I plot average beliefs of subjects who won and lost the bonus for no-motive-group subjects in Figure 7.

Figure 7 shows a clear and significant difference ($p < 0.001$, Wilcoxon rank-sum test) in no-motive-group subjects' beliefs between those who won the bonus and those who lost it. This difference is equal to 13 percentage points and is somewhat less than half as large as the effect of receiving an incentivizing signal.[30] Overall, this indicates that experiencing a win induces a general tendency to believe that one's effort is likely to be rewarded.

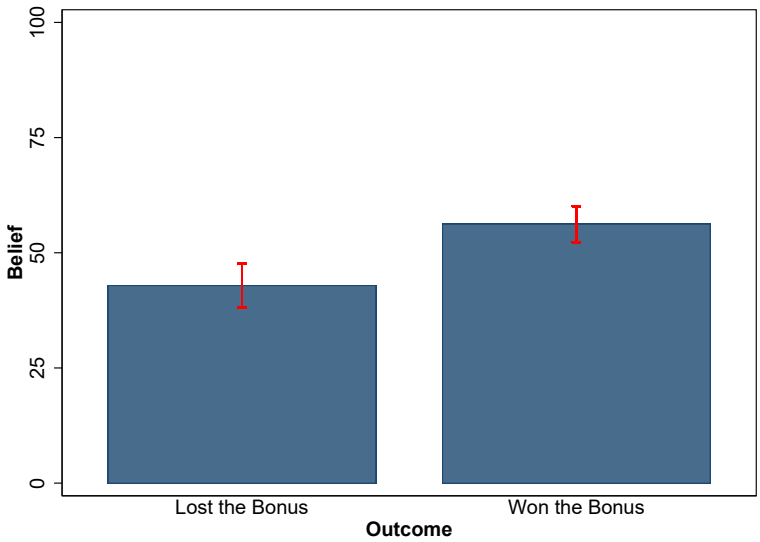
The next step is to ask whether the effect of winning the bonus on beliefs interacts in a meaningful way with the performance rank subjects hold. Figure 4, introduced in Section 5.1,

[28]Note that I can not reject the null hypothesis of no difference in the absolute *magnitude* of the interaction terms ($p = 0.453$, F-test). This means that the difference in updating is driven by a difference in direction rather than through a difference in magnitude. The reader should, however, take this null result with a grain of salt given the large standard errors and given that the magnitude of the parameters indicates that the magnitude of the distortion is larger for those subjects who received a disincentivizing signal.

[29]In the pre-analysis plan, I listed the first point under exploratory analysis. I added this part to the section on who engages in motivating belief distortion because it informs whether subjects who win or lose are more likely to engage in motivating belief distortion.

[30]Note that these average differences do not reflect the fact that winning the bonus is more informative than losing the bonus. While, for no-motive-group subjects, there is a weak correlation (0.13) between receiving an incentivizing signal and winning the bonus, this does not explain such a large difference in beliefs. This is also confirmed in Table B7 in the appendix, which reports results from linear regressions that explicitly control for the informational content of the signal. These regressions also control for task 1 performance. The coefficient of winning the bonus remains nearly unchanged, indicating that the effect from winning the bonus is unlikely to reflect unobserved heterogeneity that is correlated with ability.

Figure 7: Mean beliefs by winning and losing the bonus (no-motive-group subjects)



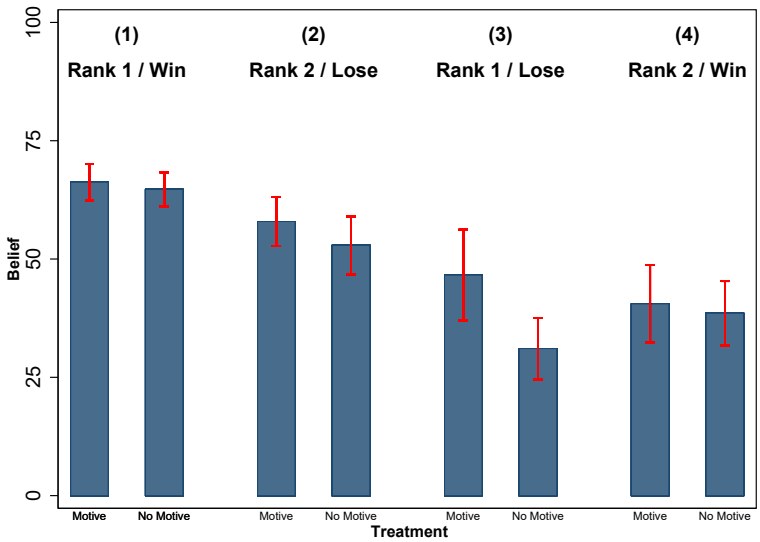
plots no-motive-group subjects' mean beliefs by winning and losing as well as rank. Though the informational content of the signal is reflected in the belief differences among those ranked second, the shift in beliefs due to the information is smaller. This can reflect subjects either over-responding to an incentivizing signal when being ranked first or under-responding to the same information when being ranked second. Though I do not have a suitable control group to definitely answer this question, I can say that subjects who ranked first and won or lost the bonus are relatively close to the Bayesian benchmark compared to those who ranked second and won or lost the bonus. The latters' beliefs are more condensed and the average beliefs of those who lost the bonus and rank second are at 55.7%, which is only slightly larger than 50% ($p = 0.005$, t-test). Hence, at the baseline, the main effect of winning or losing the bonus is likely to be driven by no-motive-group subjects under-responding to the information if they ranked second rather than first.

I now ask to what extent this effect interacts with being a motive-group subject. Figure 8 plots mean beliefs for each treatment group by every possible outcome of the task. Bars (1) and (2) feature outcomes where the informational value is incentivizing, while bars (3) and (4) feature outcomes where the motivational value is disincentivizing.

The figure clearly shows that the treatment effect is mainly driven by subjects who ranked first and lost the bonus, while there is no significant difference across treatment groups for subjects who ranked second but won the bonus. Among subjects included in bars (3), motive-group subjects are 15 percentage points more confident of being in the Performance-Environment compared to no-motive-group subjects ($p = 0.014$, Wilcoxon rank-sum test). These results were confirmed using linear regressions that control for observable heterogeneity and the informational content of the signal as presented in Table B9 in the Appendix.

This result provides evidence indicative of what type of event triggers motivating belief distortion. While I showed above that losing the bonus has a negative effect on beliefs at the baseline, this is not the case for motive-group subjects who received a disincentivizing signal.

Figure 8: Mean beliefs by winning and losing the bonus, rank, and motive treatment



This leads to the question of what type of individuals frequently observe such events outside the laboratory. Generally speaking, these are people who believe that, in a world that rewards effort, they should have won, despite the fact that they do not succeed. This type of event occurs if individuals think their productivity in a given task is above average. My results imply that, at least in the aggregate, these individuals, when facing such an event, do not update their beliefs about the state of the world to increase the role of luck. Furthermore, this is precisely the type of event that individuals face if they are persistently overconfident in their relative skill and ability. Such individuals would make relatively correct inferences about the true state of the world, holding their belief about their relative ability constant, if they could use beliefs as a motivating device. However, if they can motivate future effort by distorting their beliefs, they become significantly more confident that they are in a world that rewards effort. This equally implies that an attribution bias in luck-effort beliefs (e.g. Deffains et al., 2016) may be crowded out if beliefs can be distorted to motivate future effort.

Individual characteristics and motivating belief distortion While the above section showed how the treatment effect interacts with the type of signal, one can also ask whether certain types of individuals are more or less likely to engage in motivating belief distortion. The theoretical literature on the topic argues that individuals who are present-biased distort beliefs to overcome limited willpower (see Bénabou, 2015). Furthermore, subjects who are not very good at the task should be more likely to engage in motivating belief distortion because they need the extra motivation to perform better. Taking this as a starting point, I pre-specified to analyze whether my treatment effect interacts with work-phase-1 performance, demand to commit to the exertion of effort, and self-reported enjoyment of the task.

Table 4 shows the results from regressions that interact the motive-group dummy variable with the performance decile a subject occupied. Column (1) performs this regression for the whole sample while columns (2) and (3) split the sample by subjects receiving an incentivizing

Table 4: Regressions interacting the treatment variable with performance decile in Task 1

	(1) Belief	(2) Belief	(3) Belief
Motive group	13.45***	42.50***	2.351
	(4.975)	(12.68)	(5.365)
Bonus win	12.52***	9.391*	10.11**
	(3.239)	(5.142)	(4.563)
Task 1 perf. decile	0.831	0.756	0.543
	(0.578)	(1.000)	(0.799)
Motive * bonus win	-8.675*	-26.68***	-6.668
	(4.664)	(9.153)	(5.854)
Motive * task 1 perf. decile	-0.412	-3.905**	0.934
	(0.830)	(1.643)	(1.019)
Constant	38.82***	26.36***	50.86***
	(3.717)	(7.090)	(4.136)
Signal	All	Disincentivizing	Incentivizing
Observations	500	177	323

Note. The above table reports OLS regressions that interact the motive treatment variable with winning the bonus and performance decile in task 1. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received a disincentivizing signal and column (3) reports results for a sample that only includes subjects who received an incentivizing signal. See Table B10 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels*: *10%, **5%, ***1%.

or disincentivizing signal. All of the regressions control for the interaction effect of winning the bonus and being in the motive group. This control variable is crucial because the two effects may cancel each other out given the positive correlation between winning the bonus and performing well in the task that I discussed in the previous section. Furthermore, I showed above that the treatment effect is particularly pronounced for subjects who lost the bonus and ranked first. Hence, one must also control for the interaction effect of winning the bonus and being in the motive-group to identify an interaction effect for task 1 performance and the motive-treatment. For specifications that control for individual heterogeneity, see Table B10 in the Appendix.

The results are in line with the hypothesis that subjects who do not perform well on the task are more likely to engage in motivating belief distortion ($p = 0.019$) if they received a discouraging signal. This is consistent with the predictions from my framework, where I showed that individuals with a relatively high cost of effort are expected to distort beliefs, as they demand extra motivation to attain the threshold. Subjects who can attain the threshold with ease do not necessarily need to distort beliefs to motivate their effort and, hence, are less willing to pay the cost associated with belief distortion.[31]

[31]This result goes in some sense against the finding by Banerjee et al. (2020) who study confidence spillovers across tasks. Their heterogeneity analysis does not suggest that belief distortion is more prevalent for those who should exhibit a higher demand for motivating belief distortion. I emphasize that I view these results as complementary rather than contradictory. Importantly, the setting they study is very different than mine because

The second characteristic I examine is whether subjects who make a revealed choice to commit to effort in a future transcription task are more likely to respond to the treatment. The idea is that these are people who have a general tendency to use a commitment device to motivate future effort. As described in Section 3.7, I asked subjects to state how many images they planned to transcribe if they were reinvited two weeks hence to redo the same task under piece-rate incentives. The subjects subsequently were asked to state a minimum production level. If they transcribe fewer images than their minimum production level, they receive no pay; if they transcribe more images, they will be paid according to the piece rate. I pre-specified that a subject demands a commitment device if her minimum production level is at least half as large as the number of images she plans on transcribing. Using this binary measure yields enough variation to identify a potential heterogeneity in the treatment effect as exactly 50% of the subjects chose to commit to exert effort in the future by this measure. Furthermore, the commitment measure is uncorrelated with task 1 performance ($\rho = 0.06$, $p = 0.901$) and weakly correlated with having won the bonus after the first part ($\rho = 0.07$, $p = 0.107$).

Table 5: Regressions interacting the treatment variable with demanding a commitment device

	(1) Belief	(2) Belief	(3) Belief
Motive group	11.73***	15.10***	5.254
	(3.249)	(5.528)	(3.454)
Demands commitment	7.160**	9.578**	2.775
	(3.186)	(4.641)	(3.496)
Motive * demands commitment	-10.37**	-12.58	-5.038
	(4.503)	(7.813)	(4.792)
Constant	46.29***	30.18***	58.16***
	(2.278)	(3.206)	(2.403)
Signal	All	Disincentivizing	Incentivizing
Observations	500	177	323

Note. The above table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates the demand for a commitment device. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received a disincentivizing signal and column (3) reports results for a sample that only includes subjects who received an incentivizing signal. See Table B10 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels*: *10%, **5%, ***1%.

Table 5 presents the results of the regression and Table B11 in the Appendix replicates these regressions, controlling for observable characteristics. As before, I run the regression for the whole sample and I additionally split the sample by signal type. The results indicate that subjects who demand a commitment device are, generally speaking, more likely to believe that their effort is rewarded. The coefficients for the interaction term indicate, if anything, that subjects who demand a commitment device are *less* likely to engage in motivating belief distortion than are those who do not commit themselves to exert effort on the additional task.

their belief is ego-relevant (relative performance) and anticipatory utility motives—which I control for in my setting—may be relevant in their setting.

This goes against the pre-specified hypothesis that subjects who have a revealed preference for commitment engage in motivating belief distortion.

Finally, I test whether subjects who reported enjoying the task are less likely to engage in motivating belief distortion. The idea is that these subjects have a higher cost of effort and have low intrinsic motivation to undertake the task. Hence, it is precisely they who should inflate extrinsic incentives to motivate future effort. My findings indicate that the latter mechanism may be more relevant in this setting.

Table 6: Regressions interacting the treatment variable with enjoying the task

	(1) Belief	(2) Belief	(3) Belief
Motive group	5.779	-2.007	4.660
	(5.715)	(10.85)	(6.317)
Bonus win	13.31***	6.421	8.412**
	(3.239)	(5.095)	(4.075)
Enjoys task	-4.033	-5.876	0.889
	(4.005)	(6.856)	(4.603)
Motive * bonus win	-10.39**	-16.18**	-4.063
	(4.502)	(7.781)	(4.786)
Motive * enjoys task	7.922	22.29**	1.358
	(5.907)	(11.07)	(6.208)
Task 1 perf. decile	0.628	-0.720	1.033**
	(0.419)	(0.809)	(0.507)
Constant	42.70***	40.07***	48.43***
	(4.274)	(7.704)	(4.968)
Signal	All	Disincentivizing	Incentivizing
Observations	498	177	321

Note. The above table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject won the bonus and whether they subject enjoyed the task. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received an incentivizing signal and column (3) reports results for a sample that only includes subjects who received a disincentivizing signal. See Table B12 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels*: *10%, **5%, ***1%.

Table 6 presents the results of regressions that interact the treatment variable with a dichotomized variable indicating whether the individual enjoyed the task or not. Table B12 in the Appendix replicates this table controlling for observable heterogeneity.[32] The results are, if anything, contrary to my expectation. Subjects who enjoyed the task are more likely to engage in motivating belief distortion, as indicated by the positive and significant interaction term for subjects who received a disincentivizing signal. This result may be attributable to the fact that

[32] I additionally control for winning the bonus and its interaction because one may expect that subjects who won the bonus also feel more satisfied with undertaking the activity, which may cancel out an interaction effect with enjoying the task.

subjects who distort beliefs to motivate future effort are also prone to telling themselves that they like the task for the same reason. This, however, needs further study in greater detail in future research.

What do these results imply about the relevant mechanism behind belief distortion? The first finding outlined in this section is consistent with the framework proposed above, because one would expect that those for whom the task is hard are more inclined to pay the mental cost of belief distortion because their expected returns from it are higher. This finding is consistent with a mechanism where beliefs are used as a commitment device because it helps those subjects who need the extra motivation to get over the threshold that need the extra motivation. The result regarding the commitment device is, however, less consistent with the idea that subjects use beliefs as a commitment device because one would expect that motivating belief distorters would use similar instruments to commit themselves to exert effort. Hence, it is more likely that an alternative mechanism is relevant in this context. One alternative mechanism that was outlined in Section 2 is that subjects have an aversion to engage in an activity when they believe that the activity is unnecessary or does not serve any purpose. To avoid this extra cost of effort, individuals may distort information that indicates that their effort is not needed to obtain the reward. My results suggest that this mechanism may be more relevant in this setting than the idea that belief distortion is driven by present-biased preferences, but more research is warranted to reach a conclusive answer to this question.

Memory To test whether the difference in beliefs is explained by motive-group subjects being more likely to forget or recode the signal they previously received, I asked subjects to recall the signal at the beginning of the post-experimental questionnaire. Overall, subjects were relatively good at recalling their signal, as 86% of subjects correctly recalled the signal. There is no significant difference across treatment groups in the propensity to recall the correct signal ($p = 0.365$, Fisher's exact test). Even though subjects were 15 percentage points more likely to not recall the correct signal if they received a disincentivizing signal ($p < 0.001$, Fisher's exact test), I cannot reject the null hypothesis that motive-group-subjects are equally likely to recall the correct signal conditional on having observed a disincentivizing signal ($p = 0.375$, Fisher's exact test) or incentivizing signal ($p = 0.843$, Fisher's exact test). These results are confirmed in Table B13 using linear regressions and controlling for observable heterogeneity.

Furthermore, I can ask whether motive-group subjects are more likely to recall the reverse signal. Among those who said that they recalled a signal, as opposed to saying that they did not recall any signal, I can construct the signal they claimed to recall and construct a variable which I call "perceived signal." Recalling an incorrect signal is relatively rare and only occured for 22 subjects (5% of all subjects). Unsurprisingly, I cannot detect any group differences in the propensity to recall a false signal ($p = 0.196$, Fisher's exact test). This holds equally for subjects who received an incentivizing signal ($p = 0.35$, Fisher's exact test) or a disincentivizing signal ($p = 0.249$, Fisher's exact test). One can equally ask whether the motive-treatment effect is mediated by recalling a false signal. To that end, I regress beliefs on the signal they recall (conditional on claiming that they recalled any signal). As shown in Table B14 in the

Appendix, this does not explain the difference across treatment groups and the reverse seems to be true—i.e., the magnitude of the treatment effect becomes larger once we control for the recalled signal. These results indicate that the cognitive mechanism that yields the difference across treatment groups is unlikely to be amnesia but, rather, under-investment in decoding the information contained in the signal.

5.4 Distributive Behavior

The previous section characterized updating behavior across treatment groups. We now ask whether this difference in updating behavior is reflected in the distributive decisions subjects make. Note that this decision does not affect their own payoff and isolates *other*-regarding motives in the decision to distribute money among two agents.

Figure 9: Redistribution by motive treatment

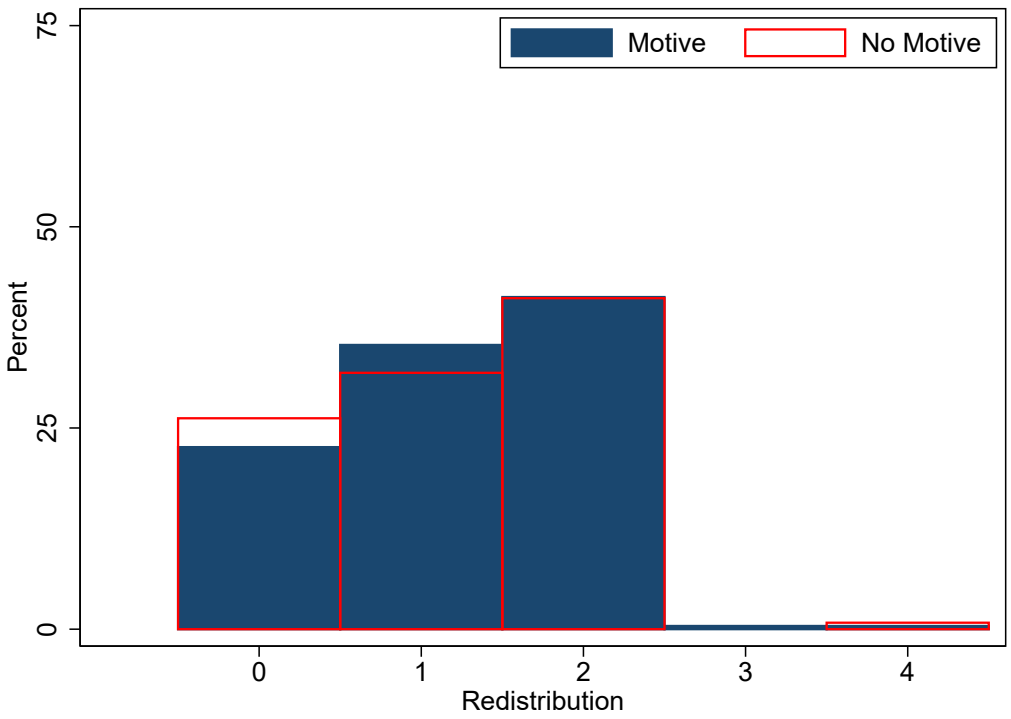


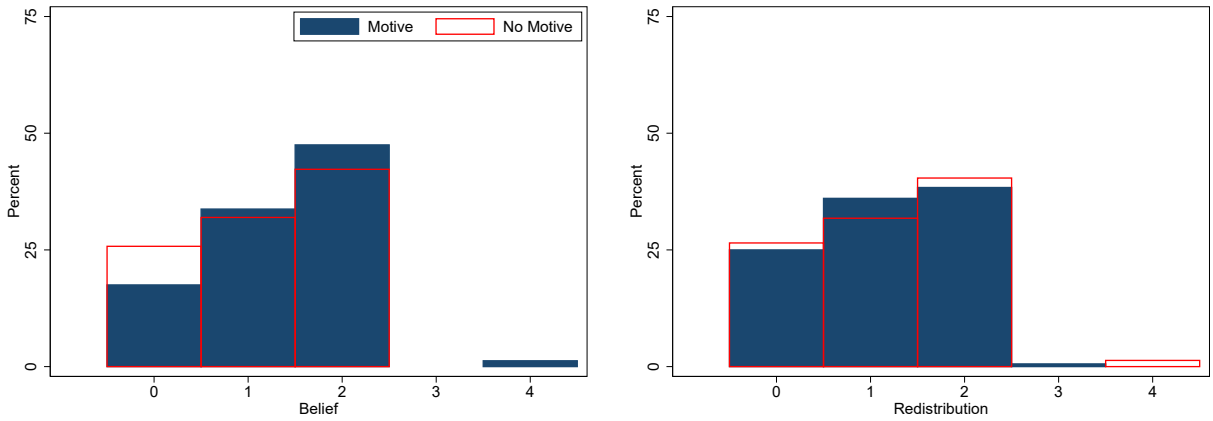
Figure 9 plots the distribution of redistributive decisions for motive- and no-motive-group subjects. A subject who redistributes 0 chooses to distribute €4 to the worker who won the initial bonus and to not give any bonus to the other worker. A subject who redistributes 2 chooses an equal split and redistributes half of the bonus to the worker who did not win the initial bonus. The graph shows that there is heterogeneity in distributive behavior. More than half of the subjects (58%) chose not to fully redistribute the initial distribution. 25% chose not to redistribute a positive amount and, instead, chose the initial distribution of income. Giving *more* to Worker B is very rare, as only 4 out of 500 subjects chose to reverse the initial inequality.

I cannot reject the null hypothesis that motive-group subjects redistribute the same amount

of money as no-motive-group subjects ($p = 0.65$, Wilcoxon rank-sum test). If anything, the data suggest that motive-group subjects are more likely to redistribute.

Figure 10: Redistribution by motive treatment and signal

(a) Disincentivizing signal (b) Incentivizing signal



Note. These figures plot the distribution of redistribution decisions by treatment group. Figure (a) plots the distribution for subjects who received a disincentivizing signal. Figure (b) plots the distribution for subjects who received an incentivizing signal.

Figure 10 plots the distribution of the redistribution decision for subjects who received an incentivizing or a disincentivizing signal in subfigures (a) and (b), respectively. In both cases, I cannot reject the null-hypothesis that redistributive behavior is equal across the two treatment groups ($p = 0.21$ (disincentivizing signal), $p = 0.81$ (incentivizing signal), Wilcoxon rank-sum test). Again, if anything the evidence indicates that motive-group subjects redistribute *more* than no-motive-group subjects, which goes against the hypothesis specified above. These results are confirmed by linear regressions that control for observable heterogeneity as presented in Table B15 in the Appendix.

Table B15 in the Appendix also presents regressions that interact the motive-treatment dummy variable with a dichotomized variable that indicates whether the subject self-identifies as a meritocrat. I do not find any evidence that belief distortion is stronger for subjects who self-identify as a meritocrat. This is true for subjects who received a disincentivizing or an incentivizing signal.

Overall, I do not find evidence that distributive behavior differs significantly across motive- and no-motive-groups. The induced distortion in beliefs through the motive treatment may not be sufficiently strong to provoke changes in distribution behavior. Most of the literature has studied distributive behavior and the source of inequality has varied the likelihood that subjects are indeed rewarded for their effort on the *extensive* margin.[33] Hence, an average difference of 7 percentage points may be too small to translate into average differences in inequality acceptance.[34] Further, Cappelen et al. (2019) show that the relationship between beliefs on the

[33]Most experiments that study how the source of inequality affects preferences regarding redistribution compare treatments where there is no uncertainty in how the initial allocation of income was distributed.

[34]Results from a pilot-experiment (N=62) show that individuals are indeed concerned about merit, if there

intensive margin and distributive behavior is convex in the probability that effort is rewarded. As mentioned above, belief distortion is particularly relevant for those who received a disincentivizing signal. Average belief probabilities (of being in the Performance-Environment) across the two treatment groups are 34% and 43%. While this variation is large in magnitude, it may not be around the point that is locally relevant to trigger differences in redistributive behavior. Future research should explore further whether this is actually the case. One could, for example, replicate the experiment but vary the prior that subjects hold. If inequality–acceptance and beliefs follow a convex relationship, one could imagine that a treatment may potentially be stronger around a 90% prior.

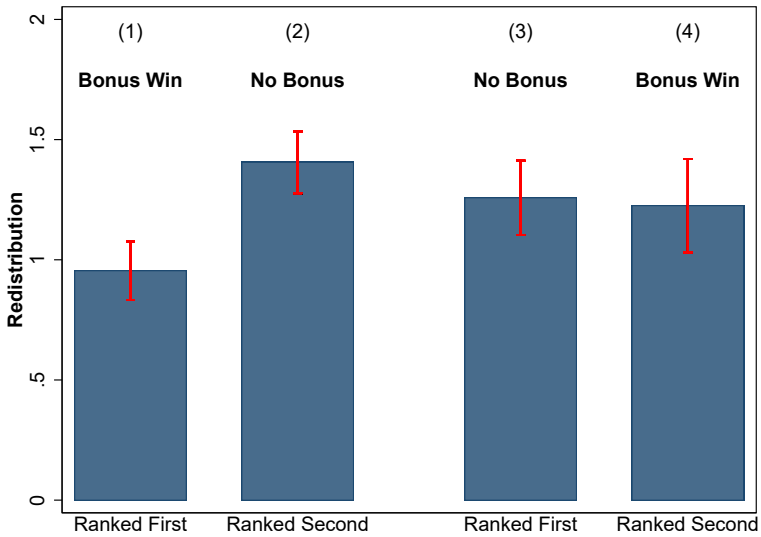
Characterizing redistributive behavior Though the motive-treatment does not directly affect distributive behavior, I do find that beliefs are significantly correlated with distributive behavior ($p = 0.041$), as shown in Table B16 in the Appendix. This dynamic is largely driven by distributive behavior at the extensive margin: Beliefs are uncorrelated with implementing complete equality ($p = 0.445$), but subjects who believe with greater certainty that they are in a Performance-Environment are significantly more likely to not redistribute at all ($p = 0.005$).[35]

One can also ask how distributive behavior varies across events. Figure 11 plots average amounts redistributed by performance rank and prize received. Bars (1) and (2) show average amounts redistributed for subjects who received an incentivizing signal and bars (3) and (4) show average amounts redistributed for subjects who received a disincentivizing signal. The figure clearly shows that subjects who won the bonus and ranked first redistributed significantly less than all other subjects ($p < 0.001$, Wilcoxon rank-sum test). Beyond that, it is apparent that subjects who received a disincentivizing signal redistribute relatively similar amounts no matter whether they won the bonus and ranked second or whether they lost the bonus and ranked first in terms of performance. On the other hand, I find that subjects who ranked second and did not win the bonus redistribute *more* than the other subjects that received an incentivizing signal ($p < 0.001$, Wilcoxon rank-sum test) and weakly more than do subjects who received a disincentivizing signal ($p = 0.081$, Wilcoxon rank-sum test). Overall, this indicates that subjects who did not win the bonus and ranked second are not making redistributive decisions based on

is certainty about the true state of the world. In that design, subjects were informed that Worker A was the better performer with 100% probability if they were in the Performance-Environment and with 50% probability if they were in the Chance-Environment. Hence, there was considerably less uncertainty about who was the better performer conditional on being in the Performance-Environment. Furthermore, it included a strategy-method decision, where subjects made the identical distribution decision *conditional* on being in the Chance- or Performance-Environment that shows that subjects redistribute 40 cents less ($p < 0.001$, t-test) if they made the decision conditional on being in the Performance-Environment compared to when they made the same decision conditional on being in the Chance-Environment. Beyond that, the correlation between beliefs and redistribution in the unconditional decision is stronger than in the main treatment ($\beta = -0.007$, $p = 0.093$). Nonetheless, I still fail to find a significant negative relationship between redistribution and being in the motive group. If anything, it is positive (diff.=0.27, $p = 0.301$, Wilcoxon rank-sum test). This is not explained by a difference in the treatment effect, which is similar in size (diff. $= 4.9$, $p = 0.386$, Wilcoxon rank-sum test) but insignificant given the small sample.

[35]These correlations should be viewed with caution because I cannot reject the null-hypothesis that beliefs are uncorrelated with distributive behavior if I instrument beliefs by the signal subjects received ($p = 0.284$). This may not be very surprising given that the signal predicts a 20-percentage-point shift in beliefs around 50%. Beyond that, the extensive margin regressions were not pre-registered. I present them because they provide a more comprehensive impression of distributive behavior.

Figure 11: Redistribution by event



their induced beliefs. If we compare, however, distributive behavior across subjects who ranked first, I find that subjects who won the bonus distribute 30 cents less than subjects who did not win the bonus ($p = 0.003$, Wilcoxon rank-sum test) and they are 15 percentage points more likely to not redistribute at all ($p = 0.005$, Fisher's exact test). Regressions reported on Table B17 confirm these results after controlling for observable heterogeneity.

Concluding this section, I do not find evidence that distributive behavior is significantly affected by the motive treatment in my context. As mentioned above, the correlation between beliefs and distributive behavior is rather weak. This stems from the fact that redistributive behavior is also heavily affected by factors that are not directly related to beliefs, such as winning the prize. This makes it challenging to identify a significant relationship between motivating belief distortion and redistributive behavior.

Taking a step back, the results complement the finding by Cassar and Klein (2019) who show that experiencing failure or success has a pure effect on distributive decisions that cannot be explained through variation in beliefs. My result shows that this effect can dominate if it comes through failure and through negative information about one's own relative ability. This may spark a feeling of solidarity with the worker who was not allocated the initial bonus. Future research should further study how these experiences shape preferences concerning redistribution above and beyond the belief channel.

6 Conclusion

This paper asks whether individuals distort their beliefs about the importance of effort in economic outcomes in order to motivate themselves to exert effort. I employ a novel experimental design where individuals receive a noisy signal about the true importance of effort in success. To identify motivating belief distortion, I vary the knowledge about a subsequent task across sub-

jects. I find that subjects distort luck-effort beliefs to motivate future effort. Subjects who know that they face a task in the future believe that it is more likely that they are in an environment where their effort is likely to be rewarded. This form of belief distortion is more pronounced for subjects who received a signal that indicates that it is rather unlikely that the true state of the world rewards effort.

My main result enhances our understanding of how individuals form luck-effort beliefs. I advance this literature by providing causal evidence that these beliefs are not only shaped by past experiences and current information about income inequality but also by the incentives individuals expect to face in the future. This has wide-ranging implications: It implies that individual beliefs about the relative importance of luck and effort may be inelastic to information about the shape of the income distribution or the degree of intergenerational mobility if individuals believe that overcoming a lack of willpower remains important (e.g. due to a lack of social security). This implies that individual beliefs about the relative importance of luck and effort depend on expected levels of redistribution and post-tax inequality. Hence, average beliefs about the importance of effort and luck may remain stable for a relatively long time, unless the electorate does not expect a shift in distributive policy in the near future. These beliefs and preferences for redistribution, which have been shown to be tightly connected, can explain why Americans' support for redistribution has been rather stable since the 1970's even though there has been a sharp increase in economic inequality over the past decades (Ashok et al., 2015).

My findings open up new avenues for research. First, it would be interesting to replicate this study while varying the prior likelihood of being in the Performance-Environment. One could expect a tighter correlation between changes in beliefs due to motivating belief distortion and preferences for redistribution if subjects hold a high prior. This would give us a more granular view of how motivating belief distortion affects preferences for redistribution. Second, it would be equally interesting to test how the magnitude of motivating belief distortion interacts with the shape of the pre-tax income distribution. The demand to distort beliefs should be higher if pre-tax inequality is expected to be high and if a significant share of low-skilled individuals are overconfident in their skill and ability. Both extensions would enable us to identify to what extent motivating belief distortion has a reinforcing effect on inequality acceptance, i.e. individuals who expect low levels of redistribution and high levels of post-tax inequality distort beliefs to motivate future effort, and this makes them less likely to vote for more redistribution, fostering inequality.

References

Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution: US vs. Europe. *American Economic Review*, 95, 2005.

Alesina, Alberto and Paola Giuliano. Chapter 4 - Preferences for Redistribution. In Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 93–131. North-Holland, 2011.

Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931, 2005.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.

Ambuehl, Sandro. An Offer You Can't Refuse? Incentives Change How We Inform Ourselves and What We Believe. *Working Paper*, 2017.

Ashok, Vevikin, Ilyana Kuziemko, and Ebonya Washington. Support for redistribution in an age of rising inequality: New stylized facts and some tentative explanations. *Brookings Papers on Economic Activity*, pages 367–405, 2015.

Augenblick, Ned and Matthew Rabin. An Experiment on Time Preference and Misprediction in Unpleasant Tasks. *The Review of Economic Studies*, 86(3):941–975, 2019.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. Working over Time: Dynamic Inconsistency in Real Effort Task. *The Quarterly Journal of Economics*, 130(3):1067–1115, 2015.

Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval. Feedback spillovers across tasks, self-confidence and competitiveness. *Games and Economic Behavior*, 123:127–170, 2020.

Barron, Kai. Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*, 2020.

Bénabou, Roland. The Economics of Motivated Beliefs. *Revue d'economie politique*, Vol. 125 (5):665–685, 2015.

Bénabou, Roland and Jean Tirole. Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 2002.

Bénabou, Roland and Jean Tirole. Belief in a Just World and Redistributive Politics. *The Quarterly Journal of Economics*, 121(2):699–746, 2006.

Bénabou, Roland and Jean Tirole. Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics*, 126(2):805–855, 2011.

Benjamin, Daniel J. Chapter 2 - Errors in probabilistic reasoning and judgment biases. In Bernheim, B. Douglas, Stefano DellaVigna, and David Laibson, editors, *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 69–186. North-Holland, 2019.

Berglas, Steven and Edward E. Jones. Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, 36(4):405–417, 1978.

Bracha, Anat and Donald J. Brown. Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75(1):67–80, 2012.

Brunnermeier, Markus K. and Jonathan A. Parker. Optimal Expectations. *American Economic Review*, 95(4):1092–1118, 2005.

Camerer, Colin and Dan Lovallo. Overconfidence and Excess Entry: An Experimental Approach. *The American Economic Review*, 89(1):306–318, 1999.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97 (3):818–827, 2007.

Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just luck: An experimental study of risk-taking and fairness. *The American Economic Review*, 103(4): 1398–1413, 2013.

Cappelen, Alexander W, Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. A Meritocratic Origin of Egalitarian Behavior. *Working Paper*, 2019.

Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden. Fair and Unfair Income Inequality. In Zimmermann, Klaus F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–25. Springer International Publishing, Cham, 2020.

Carrillo, Juan D. and Thomas Mariotti. Strategic Ignorance as a Self-Disciplining Device. *The Review of Economic Studies*, 67(3):529–544, 2000.

Cassar, Lea and Arnd H. Klein. A Matter of Perspective: How Failure Shapes Distributive Preferences. *Management Science*, 65(11):5050–5064, 2019.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven. Self-confidence and strategic behavior. *Experimental Economics*, 21(1):72–98, 2018.

Chen, Daniel L., Martin Schonger, and Chris Wickens. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

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

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao. Motivated False Memory. *Journal of Political Economy*, page 709971, 2020.

Coutts, Alexander. Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395, 2019a.

Coutts, Alexander. Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113:549–565, 2019b.

Danz, David, Lise Vesterlund, and Alistair J. Wilson. Belief Elicitation: Limiting Truth Telling with Information on Incentives. *Working Paper*, 2020.

Deffains, Bruno, Romain Espinosa, and Christian Thöni. Political self-serving bias and redistribution. *Journal of Public Economics*, 134:67–74, 2016.

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism. *American Economic Review*, 105(11):3416–3442, 2015.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550, 2011.

Durante, Ruben, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.

Erkal, Nisvan, Lata Gangadharan, and Boon Han Koh. By Chance or by Choice? Biased Attribution of Others' Outcomes when Social Preferences Matter. *Working Paper*, 2021.

Fehr, Dietmar, Hannes Rau, Stefan T. Trautmann, and Yilong Xu. Inequality, fairness and social capital. *European Economic Review*, 129:103566, 2020.

Fong, Christina. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246, 2001.

Gärtner, Manja, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution: Perceptions and Reality. *Working Paper*, 2019.

Huck, Steffen, Nora Szech, and Lukas M Wenner. More Effort with Less Pay: On Information Avoidance, Optimistic Beliefs, and Performance. *Working Paper*, page 39, 2018.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. Self-Control at Work. *Journal of Political Economy*, 123(6):1227–1277, 2015.

König, Tobias, Sebastian Schweighofer-Kodritsch, and Georg Weizsäcker. Beliefs as a Means of Self-Control? Evidence from a Dynamic Student Survey. *Working Paper*, 2019.

Konow, James. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091, 2000.

Krawczyk, Micha\l. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, 94(1):131–141, 2010.

Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143, 2016.

Lerner, Melvin J. The Belief in a Just World. In Lerner, Melvin J., editor, *The Belief in a Just World: A Fundamental Delusion*, Perspectives in Social Psychology, pages 9–30. Springer US, Boston, MA, 1980.

Möbius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat. Managing Self-Confidence. *Working Paper*, 2014.

Palan, Stefan and Christian Schitter. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

Piketty, Thomas. Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):551–584, 1995.

Schwardmann, Peter and Joël van der Weele. Deception and self-deception. *Nature Human Behaviour*, 3(10):1055–1061, 2019.

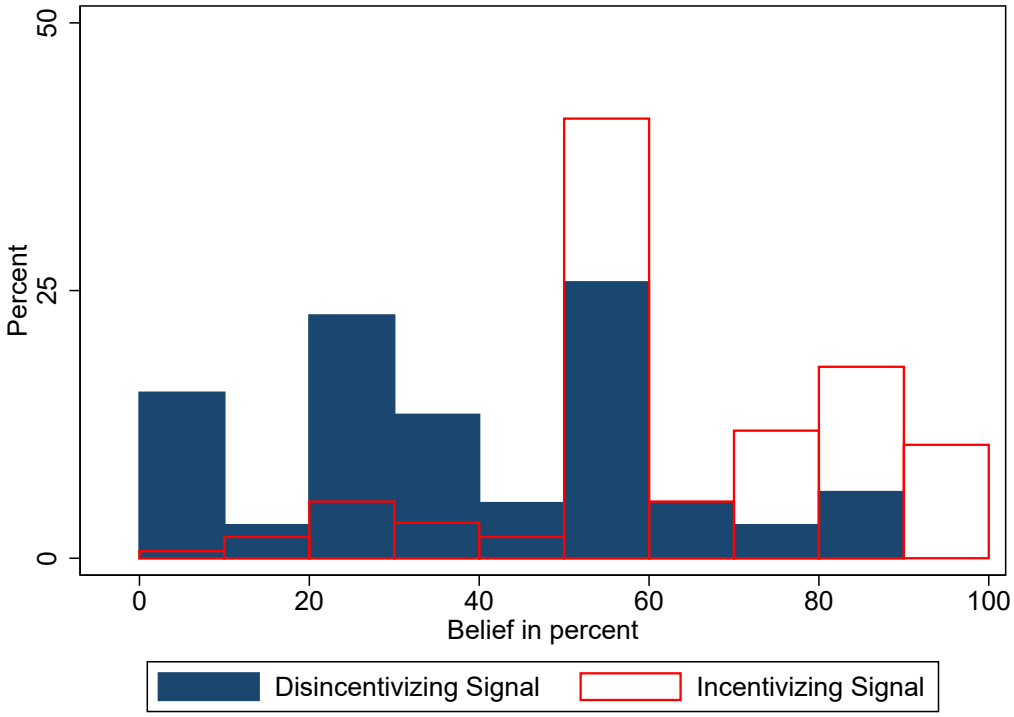
Valero, Vanessa. Redistribution and Beliefs about the Source of Income Inequality. *Working Paper*, 2020.

Vischer, Thomas, Thomas Dohmen, Armin Falk, David Huffman, Jürgen Schupp, Uwe Sunde, and Gert G. Wagner. Validating an ultra-short survey measure of patience. *Economics Letters*, 120(2):142–145, 2013.

Zimmermann, Florian. The Dynamics of Motivated Beliefs. *American Economic Review*, 110 (2):337–363, 2020.

A Figures

Figure A1: Distribution of no-motive-group subjects' beliefs by signal



B Tables

Table B1: Summary statistics

	Mean	SD	Min	Max
Motive group	0.505	0.500	0	1
Age	33.808	12.380	18	78
Female	0.406	0.492	0	1
Employed full time	0.432	0.496	0	1
Student	0.244	0.430	0	1
Risk aversion	5.180	2.261	0	10
Patience	6.119	2.307	0	10
Right wing	0.246	0.431	0	1
Observations	495			

Note. This table presents mean, standard deviation, smallest observation, and largest observation for my sample. Motive group is a dummy, indicating whether the subject was assigned ot the motive-group; age is the self-declared age; female is the a dummy indicating whether the subject self-identifies as a female; employed full-time is a dummy indicating whether the subject works full time; student is a dummy indicating whether the subject is currently a student; risk-aversion is a variable indicating the subject's attitude towards risk as in Dohmen et al. (2011); patience is a variable indicating the subject's patience Vischer et al. (2013); and rightwing is a dummy indicating whether the subject self-identifies as having a political orientation right of the center.

Table B2: Summary statistics by motive treatment

Variable	(1) No-motive-group	(2) Motive-group	(3) Difference
Age	33.439	34.156	0.717
	(12.338)	(12.413)	(1.111)
Female	0.403	0.401	-0.002
	(0.492)	(0.491)	(0.044)
Employed full time	0.427	0.440	0.013
	(0.496)	(0.497)	(0.044)
Student	0.238	0.250	0.012
	(0.427)	(0.434)	(0.038)
Risk aversion	5.000	5.381	0.381*
	(2.241)	(2.279)	(0.202)
Patience	6.211	6.048	-0.163
	(2.323)	(2.305)	(0.207)
Right wing	0.262	0.234	-0.028
	(0.441)	(0.424)	(0.039)
Observations	248	252	500

Note. This table presents results from a balancing test across motive treatment groups. The first column presents mean of the variable for no-motive-group subjects, while the second column presents the mean of the same variable for motive-group variables; the third columns characterizes the difference between the two. Standard deviation is in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B3: Summary statistics by motive treatment

Variable	(1) Disincentivizing signal	(2) Incentivizing signal	(3) Difference
Age	33.273	34.091	0.818
	(11.937)	(12.609)	(1.161)
Female	0.390	0.409	0.019
	(0.489)	(0.492)	(0.046)
Employed full time	0.412	0.446	0.033
	(0.494)	(0.498)	(0.046)
Student	0.254	0.238	-0.016
	(0.437)	(0.427)	(0.040)
Risk aversion	4.989	5.303	0.315
	(2.369)	(2.203)	(0.212)
Patience	6.220	6.078	-0.143
	(2.249)	(2.349)	(0.217)
Right wing	0.254	0.245	-0.010
	(0.437)	(0.431)	(0.040)
Observations	177	323	500

Note. This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. The first column presents mean of the variable for subjects that received a disincentivizing signal, while the second column presents the mean of the same variable for subjects that received an incentivizing signal; the third columns characterizes the difference between the two. Standard deviation is in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B4: Summary statistics by winning the bonus

Variable	(1) Not win bonus	(2) Win bonus	(3) Difference	(4) Conditional difference
age	34.782	32.858	-1.924*	-0.350
	(12.849)	(11.838)	(1.109)	(1.052)
female	0.359	0.443	0.084*	0.081*
	(0.481)	(0.498)	(0.044)	(0.045)
employed full time	0.424	0.443	0.019	0.021
	(0.495)	(0.498)	(0.044)	(0.045)
student	0.208	0.278	0.070*	0.053
	(0.407)	(0.449)	(0.038)	(0.039)
risk aversion	5.151	5.231	0.080	0.163
	(2.279)	(2.257)	(0.203)	(0.206)
patience	6.090	6.165	0.075	0.110
	(2.422)	(2.207)	(0.207)	(0.211)
rightwing	0.245	0.251	0.006	0.018
	(0.431)	(0.434)	(0.039)	(0.039)
Observations	245	255	500	

Note. This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. Column (1) presents mean of the variable for subjects that received a disincentivizing signal, while column (2) presents the mean of the same variable for subjects that received an incentivizing signal; column (3)) characterizes the difference between the two; and column (4) is the difference conditional on work-phase1 performance decile. Standard deviation is in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B5: Summary statistics by rank

Variable	(1) Rank first	(2) Rank second	(3) Difference	(4) Conditional difference
age	30.814	37.339	6.525***	2.414*
	(10.331)	(13.613)	(1.077)	(1.272)
female	0.412	0.390	-0.021	-0.008
	(0.493)	(0.489)	(0.044)	(0.053)
employed full time	0.441	0.425	-0.016	-0.031
	(0.497)	(0.495)	(0.045)	(0.053)
student	0.265	0.219	-0.045	0.019
	(0.442)	(0.415)	(0.039)	(0.045)
risk aversion	5.099	5.303	0.203	-0.047
	(2.192)	(2.350)	(0.203)	(0.234)
patience	6.129	6.128	-0.001	-0.141
	(2.321)	(2.307)	(0.208)	(0.240)
rightwing	0.232	0.268	0.036	0.000
	(0.423)	(0.444)	(0.039)	(0.045)
Observations	272	228	500	500

Note. This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. Column (1) presents mean of the variable for subjects that received a disincentivizing signal, while column (2) presents the mean of the same variable for subjects that received an incentivizing signal; column (3)) characterizes the difference between the two; and column (4) is the difference conditional on work-phase1 performance decile. Standard deviation is in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B6: Regression of belief on winning on incentivizing signal and control variables (no-motive-group only)

	(1) Belief	(2) Belief	(3) Belief
Incentivizing signal	22.11***	24.63***	23.30***
	(2.288)	(3.113)	(3.069)
Task 1 perf. decile		0.860	0.743
		(0.579)	(0.665)
Age		-0.0566	-0.0406
		(0.151)	(0.149)
Female		-0.284	-1.102
		(2.958)	(2.935)
Employed full time		6.659**	6.166**
		(2.973)	(2.944)
Student		5.353	3.689
		(3.748)	(3.764)
Risk aversion		0.414	0.420
		(0.631)	(0.617)
Patience		-0.230	-0.399
		(0.654)	(0.647)
Right wing		-8.089**	-8.289**
		(3.546)	(3.460)
Bonus win			8.697***
			(2.992)
Ranks first			-0.807
			(3.585)
Constant	36.18***	29.70***	28.86***
	(2.036)	(8.024)	(8.012)
Education FE	No	Yes	Yes
Signal	All	All	All
Observations	500	245	245

Note. This table reports OLS regressions that regress beliefs on receiving an incentivizing signal. It only includes subjects that are in the no-motive-group. Table B18 replicates this table with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B7: Regression of belief on winning on motive group dummy and control variables

	(1) Belief	(2) Belief	(3) Belief
Motive group	6.712***	6.924***	6.478***
	(2.263)	(2.237)	(2.221)
Bonus win		7.674***	7.223***
		(2.516)	(2.519)
Ranks first		2.433	1.698
		(3.049)	(3.083)
Age			-0.0944
			(0.101)
Female			-1.988
			(2.324)
Employed full time			2.651
			(2.420)
Student			3.412
			(3.012)
Risk aversion			1.117**
			(0.529)
Patience			-0.297
			(0.505)
Right wing			-4.594*
			(2.700)
Constant	49.64***	42.29***	43.71***
	(1.611)	(2.843)	(7.002)
Education FE	No	No	Yes
Observations	500	500	495

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, a dummy variable that indicates whether the subject won or lost the bonus, and a dummy that indicates whether the subject ranked first in terms of performance. Table B19 replicates this table with subjects that passed a basic comprehension test at the end of the experiment. All regressions are run on the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B8: Regression of belief on winning on motive group dummy and control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Belief	Belief	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	9.398**	9.599**	2.731	3.610	7.189*	7.578*	2.313	3.145
	(3.930)	(3.958)	(2.391)	(2.335)	(4.106)	(4.212)	(2.277)	(2.258)
Bonus win		-0.990		6.556**		-1.497		5.057*
		(4.451)		(2.903)		(4.597)		(2.844)
Task 1 perf. decile		-0.749		1.048**		-0.506		0.768
		(0.809)		(0.504)		(0.902)		(0.512)
Age					0.153	0.124	-0.365***	-0.245**
					(0.193)	(0.209)	(0.103)	(0.106)
Female					1.954	2.249	-3.288	-3.983*
					(4.242)	(4.365)	(2.373)	(2.331)
Employed full time					-3.016	-2.913	4.877*	5.087**
					(4.032)	(4.087)	(2.495)	(2.411)
Student					4.705	4.702	3.839	3.679
					(5.307)	(5.333)	(3.287)	(3.227)
Risk aversion					1.786**	1.687*	0.0761	0.284
					(0.844)	(0.878)	(0.544)	(0.542)
Patience					-0.109	-0.127	0.216	0.117
					(0.907)	(0.906)	(0.517)	(0.512)
Right wing					-2.781	-2.607	-4.585	-4.783
					(4.812)	(4.867)	(3.040)	(2.954)
Constant	34.23***	38.51***	59.54***	50.08***	22.49**	26.80**	70.49***	59.54***
	(2.368)	(6.166)	(1.746)	(3.047)	(10.33)	(11.96)	(6.705)	(7.528)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing
Observations	177	177	323	323	176	176	319	319

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable and a dummy variable that indicates whether the subject won or lost the bonus. Columns (1), (2), (5), and (6) report results for regressions that include subjects that received a disincentivizing signal, while columns (3), (4), (7), and (8) report results for regressions that include subjects that received an incentivizing signal. Table B20 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B9: Regression beliefs on the interaction of being in the motive group and winning the bonus

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	11.52***	16.27***	5.609	10.75***	14.65**	4.193
	(3.450)	(5.807)	(3.952)	(3.458)	(5.842)	(3.871)
Bonus win	12.73***	5.174	8.503**	11.82***	5.091	6.090
	(3.177)	(4.995)	(3.996)	(3.204)	(5.315)	(3.913)
Task 1 perf. decile	0.615	-0.899	1.054**	0.491	-0.632	0.774
	(0.417)	(0.816)	(0.502)	(0.456)	(0.897)	(0.512)
Motive * bonus win	-9.098**	-14.45*	-3.666	-8.407*	-15.48*	-1.905
	(4.457)	(7.827)	(4.783)	(4.511)	(8.025)	(4.782)
Age				-0.0955	0.147	-0.243**
				(0.102)	(0.209)	(0.107)
Female				-2.080	1.985	-3.996*
				(2.320)	(4.352)	(2.339)
Employed full time				2.482	-3.425	5.061**
				(2.425)	(4.131)	(2.420)
Student				2.831	4.300	3.554
				(3.027)	(5.437)	(3.234)
Risk aversion				1.103**	1.761**	0.280
				(0.528)	(0.864)	(0.543)
Patience				-0.349	-0.120	0.0957
				(0.504)	(0.903)	(0.517)
Right wing				-4.521*	-2.486	-4.780
				(2.687)	(4.890)	(2.954)
Constant	39.87***	36.67***	48.95***	42.08***	23.76*	59.04***
	(3.163)	(6.070)	(3.491)	(7.100)	(12.25)	(7.661)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Disincentivizing
Observations	500	177	323	495	176	319

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, and the interaction of the motive treatment and winning the bonus. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B22 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B10: Regression beliefs on the interaction of being in the motive group and doing well in the task

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	13.45***	42.50***	2.351	12.92***	37.97***	3.135
	(4.975)	(12.68)	(5.365)	(4.940)	(13.16)	(5.125)
Bonus win	12.52***	9.391*	10.11**	11.59***	9.109*	6.657
	(3.239)	(5.142)	(4.563)	(3.268)	(5.477)	(4.426)
Task 1 perf. decile	0.831	0.756	0.543	0.730	0.810	0.604
	(0.578)	(1.000)	(0.799)	(0.608)	(1.050)	(0.774)
Motive * bonus win	-8.675*	-26.68***	-6.668	-7.941*	-26.31***	-2.908
	(4.664)	(9.153)	(5.854)	(4.723)	(9.411)	(5.862)
Motive * task 1 perf. decile	-0.412	-3.905**	0.934	-0.461	-3.436**	0.307
	(0.830)	(1.643)	(1.019)	(0.828)	(1.658)	(0.987)
Age				-0.0978	0.115	-0.242**
				(0.102)	(0.218)	(0.107)
Female				-2.044	1.785	-4.048*
				(2.325)	(4.363)	(2.334)
Employed full time				2.521	-3.558	5.022**
				(2.429)	(4.212)	(2.426)
Student				2.735	2.316	3.517
				(3.010)	(5.494)	(3.248)
Risk aversion				1.086**	1.629*	0.283
				(0.528)	(0.866)	(0.545)
Patience				-0.366	-0.0879	0.109
				(0.505)	(0.899)	(0.522)
Right wing				-4.518*	-1.774	-4.764
				(2.683)	(4.937)	(2.959)
Constant	38.82***	26.36***	50.86***	41.19***	16.35	59.58***
	(3.717)	(7.090)	(4.136)	(7.241)	(12.52)	(7.884)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	500	177	323	495	176	319

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, performance decile, and the interaction of the motive treatment with the latter two variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B23 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B11: Regression beliefs on the interaction of being in the motive group and demanding committing to exert future effort

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	11.73***	15.10***	5.254	11.93***	14.45**	4.200
	(3.249)	(5.528)	(3.454)	(3.208)	(5.808)	(3.263)
Demands commitment	7.160**	9.578**	2.775	8.248***	11.14**	2.204
	(3.186)	(4.641)	(3.496)	(3.171)	(5.160)	(3.330)
Motive * demands commitment	-10.37**	-12.58	-5.038	-11.62**	-15.54*	-3.777
	(4.503)	(7.813)	(4.792)	(4.528)	(8.347)	(4.691)
Age				-0.140	0.164	-0.358***
				(0.0950)	(0.186)	(0.101)
Female				-1.405	2.734	-3.410
				(2.367)	(4.382)	(2.452)
Employed full time				2.285	-3.722	4.762*
				(2.457)	(3.952)	(2.510)
Student				3.607	4.673	3.729
				(3.038)	(5.205)	(3.311)
Risk aversion				1.029*	1.708**	0.104
				(0.525)	(0.856)	(0.553)
Patience				-0.269	-0.195	0.206
				(0.512)	(0.924)	(0.522)
Right wing				-4.884*	-3.775	-4.712
				(2.745)	(4.931)	(3.057)
Constant	46.29***	30.18***	58.16***	47.29***	17.88*	69.18***
	(2.278)	(3.206)	(2.403)	(6.263)	(10.78)	(6.730)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	500	177	323	495	176	319

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, enjoying the commitment dummy, and their interaction . Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B24 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B12: Regression beliefs on the interaction of being in the motive group and enjoying the task

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	5.779	-2.007	4.660	5.959	-2.394	3.541
	(5.715)	(10.85)	(6.317)	(5.675)	(11.13)	(6.116)
Bonus win	13.31***	6.421	8.412**	12.41***	6.235	5.861
	(3.239)	(5.095)	(4.075)	(3.266)	(5.456)	(3.955)
Motive * bonus win	-10.39**	-16.18**	-4.063	-9.527**	-16.86**	-2.200
	(4.502)	(7.781)	(4.786)	(4.544)	(8.056)	(4.739)
Enjoys task	-4.033	-5.876	0.889	-3.626	-5.841	1.811
	(4.005)	(6.856)	(4.603)	(4.030)	(7.667)	(4.287)
Motive * enjoys task	7.922	22.29**	1.358	6.687	20.85*	0.941
	(5.907)	(11.07)	(6.208)	(5.856)	(11.68)	(5.823)
Task 1 perf. decile	0.628	-0.720	1.033**	0.506	-0.470	0.739
	(0.419)	(0.809)	(0.507)	(0.460)	(0.892)	(0.523)
Age				-0.107	0.137	-0.244**
				(0.103)	(0.209)	(0.106)
Female				-2.131	2.811	-3.769
				(2.331)	(4.267)	(2.352)
Employed full time				2.411	-4.001	5.180**
				(2.433)	(4.041)	(2.456)
Student				2.271	3.210	3.267
				(3.037)	(5.379)	(3.254)
Risk aversion				1.150**	1.606*	0.232
				(0.530)	(0.861)	(0.556)
Patience				-0.283	-0.0538	0.0940
				(0.509)	(0.947)	(0.518)
Right wing				-4.379	-2.089	-4.998*
				(2.698)	(4.856)	(3.000)
Constant	42.70***	40.07***	48.43***	44.62***	27.61**	58.26***
	(4.274)	(7.704)	(4.968)	(7.563)	(12.51)	(8.174)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	498	177	321	493	176	317

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, enjoying the task dummy, and the interaction of the motive treatment with the latter two dummy variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B25 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B13: Regressing false recall on motive group, signal, its interaction term and observables

	(1) False recall	(2) False recall	(3) False recall	(4) False recall
Motive group	-0.0302	-0.0190	-0.0680	-0.0610
	(0.0309)	(0.0301)	(0.0637)	(0.0638)
Incentivizing signal		-0.152***	-0.189***	-0.176***
		(0.0355)	(0.0503)	(0.0500)
Motive * pos. signal			0.0758	0.0562
			(0.0708)	(0.0721)
Task 1 perf. decile				-0.0144**
				(0.00580)
Age				-0.000346
				(0.00167)
Female				-0.0647**
				(0.0305)
Employed full time				-0.0297
				(0.0311)
Student				-0.0332
				(0.0383)
Risk aversion				0.00351
				(0.00757)
Patience				0.00128
				(0.00671)
Right wing				0.00643
				(0.0385)
Constant	0.153***	0.246***	0.268***	0.366***
	(0.0229)	(0.0357)	(0.0452)	(0.108)
Education FE	No	No	No	Yes
Signal	All	All	All	All
Observations	500	500	500	495

Note. This table reports OLS regressions that regress not recalling the correct signal on a motive group dummy, receiving an incentivizing signal, and their interaction. Not recalling the correct signal includes (i) recalling a false signal or (ii) declaring that one did not remember the signal. The sample includes subjects all subjects. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B14: Regression beliefs controlling for recalled signal

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief	(7) Belief	(8) Belief
Motive group	5.298**	6.152***	10.82**	11.49***	2.421	2.500	9.642**	2.803
	(2.211)	(2.213)	(4.226)	(4.220)	(2.532)	(2.543)	(4.456)	(2.397)
Incentivizing signal	23.82***							
	(2.454)							
Recalled signal		23.95***		10.12*		12.36***	11.81*	14.54***
		(2.545)		(5.774)		(1.742)	(6.430)	(4.299)
Task 1 perf. decile							0.00380	1.315***
							(0.863)	(0.469)
Age							0.111	-0.280**
							(0.234)	(0.112)
Female							2.892	-3.007
							(4.569)	(2.467)
Employed full time							-2.266	5.127**
							(4.180)	(2.541)
Student							7.021	4.478
							(5.631)	(3.369)
Risk aversion							1.744*	0.176
							(0.919)	(0.570)
Patience							0.194	0.204
							(0.937)	(0.554)
Right wing							-0.181	-3.195
							(5.193)	(3.218)
Constant	34.51***	32.94***	32***	30.33***	59.86***	47.50***	16.15	44.71***
	(2.198)	(2.373)	(2.560)	(2.819)	(1.850)	(2.543)	(12.72)	(9.791)
Education FE	No	No	No	No	No	No	Yes	Yes
Signal	All	All	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing
Observations	454	454	156	156	298	298	156	295

Note. This table reports OLS regressions that regress belief on the motive treatment variable. Column (1) and (2) report results from a regression that excludes subjects that did not forget the forget the signal. Incentivizing signal denotes the actual signal and recalled signal denotes the signal they claim to recall. Column (3) regresses beliefs on the treatment dummy for subject that recalled a signal for subjects that received a disincentivizing signal. Columns (4) and (7) regress beliefs on the treatment dummy controlling for the signal subjects recalled for subjects that received a disincentivizing signal. Column (5) regresses beliefs on the treatment dummy for subject that recalled a signal for subjects that received an incentivizing signal. Columns (6) and (8) regress beliefs on the treatment dummy controlling for the signal subjects *recalled* for subjects that received a disincentivizing signal. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B15: Redistribution regression interacting the treatment variable with merit dummy

	(1) Redistr.	(2) Redistr.	(3) Redistr.	(4) Redistr.	(5) Redistr.	(6) Redistr.
Motive group	0.0139	0.264	-0.110	0.000668	0.124	-0.0600
	(0.210)	(0.298)	(0.268)	(0.209)	(0.307)	(0.276)
Meritocr.	-0.317**	-0.165	-0.405**	-0.294**	-0.181	-0.402**
	(0.140)	(0.247)	(0.169)	(0.135)	(0.226)	(0.176)
Motive * meritocr-	0.0386	-0.0915	0.109	0.0578	-0.00500	0.0708
	(0.225)	(0.326)	(0.286)	(0.223)	(0.332)	(0.295)
Age				0.00292	0.00678	0.000239
				(0.00344)	(0.00602)	(0.00428)
Female				-0.234***	-0.245*	-0.205**
				(0.0812)	(0.141)	(0.100)
Employed full time				-0.0574	-0.0487	-0.0514
				(0.0828)	(0.140)	(0.103)
Student				0.00302	-0.0380	0.0256
				(0.0968)	(0.155)	(0.124)
Risk aversion				0.00798	-0.00215	0.00780
				(0.0182)	(0.0282)	(0.0238)
Patience				-0.00317	-0.0602**	0.0287
				(0.0157)	(0.0241)	(0.0205)
Right wing				-0.0576	-0.230	0.0379
				(0.0912)	(0.150)	(0.118)
Constant	1.444***	1.308***	1.522***	1.507***	1.723***	1.470***
	(0.127)	(0.230)	(0.150)	(0.204)	(0.345)	(0.255)
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Education FE	No	No	No	Yes	Yes	Yes
Observations	500	177	323	495	176	319

Note. This table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject self-identifies as a meritocrat (agrees with 7 out of 10 that subjects who transcribed more images should receive more money). Columns (1) and (4) report results for the whole sample; column (2) and (5) report results for a sample that only include subjects who received an incentivizing signal and column (3) and (6) reports results for a sample that only include subjects who received a disincentivizing signal. Table B26 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B16: Regression on extensive margin redistribution behavior

	(1) Redistr.	(2) No redistr.	(3) Full redistr	(4) Redistr.	(5) No redistr.	(6) Full redistr
Belief	-0.00295**	0.00219***	-0.000654	-0.00262*	0.00221***	-0.000317
	(0.00144)	(0.000772)	(0.000855)	(0.00145)	(0.000774)	(0.000877)
Age				0.000806	-0.000473	0.00105
				(0.00372)	(0.00201)	(0.00216)
Female				-0.243***	0.141***	-0.127***
				(0.0815)	(0.0419)	(0.0459)
Employed full time				-0.0459	0.00690	-0.0143
				(0.0825)	(0.0419)	(0.0488)
Student				0.0249	-0.0945*	-0.0410
				(0.0985)	(0.0518)	(0.0588)
Risk aversion				0.00704	0.000877	0.000901
				(0.0180)	(0.00889)	(0.0104)
Patience				-0.00655	-0.000222	-0.00740
				(0.0155)	(0.00785)	(0.00968)
Right wing				-0.0957	0.0692	-0.0592
				(0.0919)	(0.0476)	(0.0530)
Constant	1.346***	0.128***	0.447***	1.596***	0.0323	0.615***
	(0.0827)	(0.0428)	(0.0508)	(0.221)	(0.118)	(0.138)
Perform. decile FE	No	No	No	Yes	Yes	Yes
Observations	500	500	500	495	495	495

Note. This table reports OLS regression that characterize the effect of beliefs on redistribution behavior. Columns (1) and (4) regress amount redistributed on beliefs; columns (2) and (5) regress a dummy variable that indicates whether the subject has redistributed at all; columns (3) and (6) regresses a dummy variable that indicates an equal split on beliefs. These regressions are run for the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B17: Regression on extensive margin redistribution behavior

	(1) Redistr.	(2) Redistr.	(3) Redistr.	(4) Redistr.	(5) Redistr.	(6) Redistr.	(7) Redistr.	(8) Redistr.
Bonus win	-0.0285	-0.507***			0.0187	-0.467***		
	(0.156)	(0.109)			(0.150)	(0.114)		
Task 1 perf. decile	0.00141	0.0171	0.00971	0.00804	0.0196	0.00467	0.0161	0.00985
	(0.0256)	(0.0187)	(0.0160)	(0.0162)	(0.0251)	(0.0202)	(0.0170)	(0.0171)
Incentivizing signal			0.180*	-0.300***			0.160*	-0.268***
			(0.0935)	(0.0906)			(0.0947)	(0.0924)
Age					0.00924	-0.00524	0.00310	0.00260
					(0.00642)	(0.00444)	(0.00437)	(0.00460)
Female					-0.266*	-0.180*	-0.230**	-0.239**
					(0.139)	(0.0980)	(0.102)	(0.0956)
Employed full time					-0.0554	-0.0425	-0.136	0.0331
					(0.142)	(0.0997)	(0.104)	(0.0968)
Student					-0.0135	0.0389	-0.104	0.113
					(0.157)	(0.122)	(0.123)	(0.109)
Risk aversion					0.00832	-0.00338	0.0206	-0.00482
					(0.0269)	(0.0230)	(0.0211)	(0.0200)
Patience					-0.0585**	0.0266	-0.0244	-0.0131
					(0.0240)	(0.0196)	(0.0177)	(0.0193)
Right wing					-0.276*	0.0288	-0.111	-0.150
					(0.146)	(0.116)	(0.114)	(0.104)
Constant	1.248***	1.347***	1.192***	1.201***	1.391***	1.563***	1.361***	1.342***
	(0.191)	(0.0927)	(0.105)	(0.105)	(0.373)	(0.255)	(0.256)	(0.267)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Incentivizing	All	All	Disincentivizing	Incentivizing	All	All
Observations	177	323	325	352	176	319	322	349

Note. This table reports OLS regressions that characterize redistribution behavior across different events observed by subjects. Columns (1) and (5) characterizes the effect of winning the bonus for subjects that received a disincentivizing signal; columns (2) and (6) characterizes the effect of winning the bonus for subjects that received an incentivizing signal; columns (3) and (7) characterizes the effect of receiving an incentivizing signal excluding subjects that ranked first and won the bonus; columns (4) and (8) characterizes the effect of receiving an incentivizing signal excluding subjects that ranked second and did not win the bonus. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B18: Regression of belief on winning on incentivizing signal and control variables (no-motive-group only), replication

	(1) Belief	(2) Belief	(3) Belief
Incentivizing signal	23.57***	26.63***	24.85***
	(2.464)	(3.338)	(3.397)
Task 1 perf. decile		0.623	0.908
		(0.636)	(0.720)
Age		-0.104	-0.119
		(0.174)	(0.165)
Female		0.439	0.381
		(3.179)	(3.168)
Employed full time		5.411*	5.043
		(3.229)	(3.167)
Student		6.515*	4.949
		(3.707)	(3.736)
Risk aversion		0.476	0.557
		(0.665)	(0.655)
Patience		-0.0956	-0.204
		(0.759)	(0.746)
Right wing		-9.297**	-10.00***
		(3.787)	(3.636)
Bonus win			8.201**
			(3.236)
Ranks first			-4.094
			(3.945)
Constant	35.70***	29.31***	28.16***
	(2.176)	(9.109)	(8.989)
Education FE	No	Yes	Yes
Signal	All	All	All
Observations	413	198	198

Note. This table reports OLS regressions that regress beliefs on receiving an incentivizing signal. It only includes subjects that are in the no-motive-group. This table replicates Table B6 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B19: Regression of belief on winning on motive group dummy and control variables, replication

	(1) Belief	(2) Belief	(3) Belief	(4) Belief
Motive group	5.896**	6.206**	5.786**	5.987**
	(2.514)	(2.498)	(2.492)	(2.485)
Bonus win		7.917***		7.058**
		(2.758)		(2.747)
Ranks first		3.058		1.684
		(3.522)		(3.558)
Task 1 perf. decile		0.332		0.279
		(0.583)		(0.618)
Age			-0.189*	-0.131
			(0.112)	(0.119)
Female			-0.558	-1.135
			(2.594)	(2.571)
Employed full time			1.835	2.037
			(2.780)	(2.738)
Student			4.725	4.443
			(3.291)	(3.243)
Risk aversion			1.055*	1.100*
			(0.577)	(0.586)
Patience			-0.163	-0.234
			(0.595)	(0.585)
Right wing			-3.443	-3.448
			(3.197)	(3.145)
Constant	50.00***	42.02***	51.36***	43.63***
	(1.763)	(3.468)	(6.729)	(8.155)
Education FE	No	No	Yes	Yes
Signal	All	All	All	All
Observations	413	413	409	409

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, a dummy variable that indicates whether the subject won or lost the bonus, and a dummy that indicates whether the subject ranked first in terms of performance. This table replicates Table B7 with subjects that passed a basic comprehension test at the end of the experiment. All regressions are run on the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B20: Regression of belief on winning on motive group dummy and control variables, replication

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief	(7) Belief	(8) Belief
Motive group	9.389**	9.719**	2.337	3.242	7.679*	8.158*	2.024	2.722
	(4.169)	(4.185)	(2.608)	(2.590)	(4.352)	(4.436)	(2.492)	(2.504)
Bonus win		-3.065		6.287*		-3.441		3.699
		(5.267)		(3.258)		(5.503)		(3.160)
Task 1 perf. decile		-0.972		0.791		-0.699		0.517
		(0.929)		(0.575)		(1.004)		(0.592)
Age					0.154	0.143	-0.384***	-0.295**
					(0.229)	(0.253)	(0.109)	(0.116)
Female					2.777	3.050	-3.736	-4.194
					(4.612)	(4.718)	(2.568)	(2.555)
Employed full time					-3.282	-3.216	4.507*	4.930*
					(4.424)	(4.528)	(2.662)	(2.650)
Student					7.184	7.357	5.087	5.165
					(5.919)	(6.054)	(3.170)	(3.191)
Risk aversion					1.641*	1.519	0.230	0.328
					(0.913)	(0.959)	(0.605)	(0.602)
Patience					0.186	0.158	0.258	0.174
					(1.006)	(1.001)	(0.597)	(0.591)
Right wing					-3.768	-3.448	-2.830	-2.955
					(5.284)	(5.405)	(3.418)	(3.390)
Constant	33.54***	39.89***	60.66***	52.08***	19.23*	24.92*	69.35***	61.29***
	(2.516)	(7.434)	(1.854)	(3.699)	(11.23)	(13.80)	(7.186)	(8.552)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing
Observations	154	154	259	259	153	153	256	256

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable and a dummy variable that indicates whether the subject won or lost the bonus. Columns (1), (2), (5), and (6) report results for regressions that include subjects that received a disincentivizing signal, while columns (3), (4), (7), and (8) report results for regressions that include subjects that received an incentivizing signal. This table replicates Table B8 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B21: Results of the updating regression specified in (9) (subjects passed comprehension test at the end of experiment)

Coefficient	**Estimate**
α_+	1.36
	(0.244)
β_+	0.32
	(0.361)
α_-	1.14
	(0.193)
β_-	-0.71**
	(0.322)
$H_0 : \alpha_+ = \alpha_- = 1$	$F(2, 409) = 1.12$
$H_0 : \beta_+ = \beta_-$	$F(1, 409) = 3.77$
N	413
R^2	0.16

Note. This table replicates Table B21 excluding subjects that did not answer correctly a comprehension question at the end of the experiment. It shows coefficient estimates for the regression characterized in (9). α coefficients test the null-hypothesis that $\alpha = 1$ while β coefficients test the null-hypothesis that $\beta = 0$. Robust standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B22: Regression beliefs on the interaction of being in the motive group and winning the bonus (subjects passed comprehension test at the end of experiment)

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	10.96***	17.03***	3.793	10.74***	16.50***	2.230
	(3.964)	(5.941)	(4.696)	(3.941)	(5.940)	(4.503)
Bonus win	13.03***	4.568	6.764	11.95***	5.223	3.258
	(3.481)	(5.739)	(4.446)	(3.487)	(6.225)	(4.276)
Task 1 perf. decile	0.637	-1.104	0.797	0.437	-0.847	0.512
	(0.464)	(0.920)	(0.573)	(0.517)	(0.981)	(0.595)
Motive * bonus win	-8.904*	-16.41**	-0.921	-8.929*	-18.85**	0.820
	(5.008)	(8.266)	(5.491)	(5.010)	(8.551)	(5.359)
Age				-0.139	0.161	-0.296**
				(0.118)	(0.248)	(0.117)
Female				-1.044	3.393	-4.192
				(2.573)	(4.689)	(2.561)
Employed full time				1.838	-4.050	4.941*
				(2.727)	(4.562)	(2.660)
Student				3.957	7.075	5.219
				(3.213)	(6.143)	(3.220)
Risk aversion				1.129*	1.596*	0.323
				(0.579)	(0.939)	(0.604)
Patience				-0.261	0.212	0.183
				(0.584)	(0.990)	(0.596)
Right wing				-3.666	-4.059	-2.932
				(3.122)	(5.404)	(3.410)
Constant	39.39***	37.39***	51.75***	41.52***	20.78	61.56***
	(3.810)	(7.204)	(4.286)	(8.274)	(13.82)	(8.844)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, and the interaction of the motive treatment and winning the bonus. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B9 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B23: Regression beliefs on the interaction of being in the motive group and doing well in the task

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	13.94**	48.86***	-0.470	14.65**	44.69***	2.654
	(6.094)	(14.21)	(6.719)	(5.989)	(14.59)	(6.258)
Bonus win	12.83***	11.10*	8.573*	11.69***	11.20*	3.057
	(3.516)	(5.793)	(5.001)	(3.525)	(6.465)	(4.805)
Task 1 perf. decile	0.956	1.019	0.160	0.852	0.974	0.578
	(0.657)	(1.068)	(0.932)	(0.697)	(1.064)	(0.912)
Motive * bonus win	-8.507*	-32.50***	-4.368	-8.400	-32.86***	1.172
	(5.124)	(10.22)	(6.533)	(5.141)	(10.65)	(6.486)
Motive * task 1 perf. decile	-0.582	-4.637**	1.122	-0.765	-4.049**	-0.113
	(0.924)	(1.797)	(1.170)	(0.921)	(1.803)	(1.124)
Age				-0.142	0.122	-0.295**
				(0.118)	(0.252)	(0.117)
Female				-0.982	3.871	-4.175
				(2.583)	(4.710)	(2.565)
Employed full time				1.974	-4.263	4.974*
				(2.735)	(4.672)	(2.688)
Student				3.890	4.946	5.259
				(3.186)	(6.169)	(3.247)
Risk aversion				1.110*	1.357	0.326
				(0.578)	(0.943)	(0.604)
Patience				-0.285	0.335	0.177
				(0.586)	(1.001)	(0.602)
Right wing				-3.808	-3.680	-2.973
				(3.104)	(5.474)	(3.389)
Constant	37.69***	23.59***	54.43***	39.61***	10.40	61.28***
	(4.653)	(7.977)	(5.323)	(8.586)	(13.74)	(9.272)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, performance decile, and the interaction of the motive treatment with the latter two variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B10 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B24: Regression beliefs on the interaction of being in the motive group and demanding committing to exert future effort (subjects passed comprehension test at the end of experiment)

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	10.11***	12.99**	4.177	10.88***	12.92**	3.479
	(3.492)	(5.851)	(3.700)	(3.456)	(6.107)	(3.564)
Demands commitment	5.862*	5.847	2.721	7.339**	8.954	1.615
	(3.521)	(5.107)	(3.731)	(3.545)	(5.988)	(3.661)
Motive * demands commitment	-9.164*	-8.184	-3.939	-11.21**	-12.13	-3.094
	(5.024)	(8.412)	(5.216)	(5.081)	(9.074)	(5.198)
Age				-0.177	0.154	-0.381***
				(0.110)	(0.223)	(0.108)
Female				-0.720	3.314	-3.816
				(2.629)	(4.767)	(2.677)
Employed full time				1.493	-3.804	4.448*
				(2.786)	(4.415)	(2.677)
Student				4.379	6.620	4.996
				(3.267)	(5.849)	(3.191)
Risk aversion				1.142*	1.699*	0.258
				(0.583)	(0.927)	(0.616)
Patience				-0.173	0.114	0.264
				(0.596)	(1.014)	(0.602)
Right wing				-4.085	-5.444	-2.927
				(3.203)	(5.521)	(3.438)
Constant	47.43***	31.25***	59.38***	47.58***	15.82	68.44***
	(2.382)	(3.265)	(2.499)	(6.976)	(11.66)	(7.180)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, enjoying the commitment dummy, and their interaction . Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B11 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B25: Regression beliefs on the interaction of being in the motive group and enjoying the task (subjects passed comprehension test at the end of experiment)

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Motive group	2.657	-1.917	0.702	3.740	-1.121	-0.331
	(6.224)	(11.19)	(7.121)	(6.273)	(11.57)	(6.933)
Bonus win	13.76***	6.566	6.570	12.75***	7.278	2.865
	(3.548)	(5.758)	(4.509)	(3.556)	(6.324)	(4.304)
Motive * bonus win	-10.85**	-18.56**	-1.943	-10.66**	-20.75**	0.0984
	(5.052)	(8.167)	(5.434)	(5.046)	(8.578)	(5.258)
Enjoys task	-4.421	-7.924	2.354	-4.077	-8.121	3.146
	(4.269)	(7.410)	(4.906)	(4.372)	(8.128)	(4.629)
Motive * enjoys task	11.67*	23.62**	4.450	9.996	22.31*	3.530
	(6.358)	(11.45)	(6.699)	(6.466)	(12.28)	(6.494)
Task 1 perf. decile	0.635	-0.892	0.706	0.434	-0.655	0.391
	(0.466)	(0.905)	(0.577)	(0.520)	(0.967)	(0.612)
Age				-0.159	0.137	-0.308***
				(0.119)	(0.246)	(0.116)
Female				-0.941	4.288	-3.612
				(2.567)	(4.562)	(2.535)
Employed full time				1.690	-4.803	4.984*
				(2.728)	(4.456)	(2.718)
Student				3.089	5.505	4.533
				(3.205)	(6.118)	(3.213)
Risk aversion				1.153**	1.464	0.206
				(0.580)	(0.938)	(0.618)
Patience				-0.151	0.346	0.212
				(0.598)	(1.023)	(0.595)
Right wing				-3.471	-3.353	-3.369
				(3.120)	(5.355)	(3.430)
Constant	42.43***	41.88***	50.64***	44.38***	25.45*	60.96***
	(4.838)	(8.928)	(5.743)	(8.586)	(14.10)	(9.244)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	411	154	257	407	153	254

Note. This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, enjoying the task dummy, and the interaction of the motive treatment with the latter two dummy variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B25 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table B26: Redistribution regression interacting the treatment variable with merit dummy (subjects passed comprehension test at the end of experiment)

	(1) Redistr.	(2) Redistr.	(3) Redistr.	(4) Redistr.	(5) Redistr.	(6) Redistr.
Motive group	0.0552	0.264	-0.0673	0.0835	0.180	0.0509
	(0.244)	(0.299)	(0.340)	(0.236)	(0.304)	(0.340)
Meritocr.	-0.217	-0.126	-0.281	-0.180	-0.183	-0.206
	(0.163)	(0.250)	(0.215)	(0.147)	(0.227)	(0.211)
Motive * meritocr-	-0.00600	-0.137	0.0778	-0.0366	-0.0984	-0.0333
	(0.259)	(0.330)	(0.358)	(0.251)	(0.333)	(0.359)
Age				0.00636	0.00923	0.00427
				(0.00409)	(0.00679)	(0.00513)
Female				-0.205**	-0.217	-0.189*
				(0.0885)	(0.150)	(0.112)
Employed full time				-0.0927	-0.0921	-0.0602
				(0.0936)	(0.147)	(0.120)
Student				0.0800	-0.00264	0.126
				(0.113)	(0.167)	(0.151)
Risk aversion				-0.00963	-0.0127	-0.0119
				(0.0203)	(0.0303)	(0.0273)
Patience				-0.00667	-0.0571**	0.0222
				(0.0175)	(0.0256)	(0.0236)
Right wing				0.0102	-0.107	0.0798
				(0.109)	(0.169)	(0.149)
Constant	1.345***	1.308***	1.375***	1.319***	1.568***	1.220***
	(0.149)	(0.231)	(0.197)	(0.239)	(0.371)	(0.320)
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Education FE	No	No	No	Yes	Yes	Yes
Observations	413	154	259	409	153	256

Note. This table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject self-identifies as a meritocrat (agrees with 7 out of 10 that subjects who transcribed more images should receive more money). Columns (1) and (4) report results for the whole sample; column (2) and (5) report results for a sample that only include subjects who received an incentivizing signal and column (3) and (6) reports results for a sample that only include subjects who received a disincentivizing signal. This table replicates Table B15 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels*: *10%, **5%, ***1%.

C Questionnaire

- Age
- Gender
- Country of living
- Nationality
- Level of Education
- Field of study
- Political orientation
- Risk aversion (Dohmen et al., 2011)
- Patience (Vischer et al., 2013)
- Question to characterize self-reported procrastination behavior: "I tend to put things off until later, although it would be better to do them right away." Subjects then select one of the following items: *Absolutely not like me*, *Very little like me*, *Not really like me*, *A little like me*, *Very similar to me*, *Absolutely like me*, *I do not know*
- Question to characterize self-reported tendency to take measures to overcome self-control problems (situational): "I sometimes choose to not go to a gathering in the evening because if I go, I may stay longer than initially planned and end up being tired the next day." Subjects then select one of the following items: *Absolutely not like me*, *Very little like me*, *Not really like me*, *A little like me*, *Very similar to me*, *Absolutely like me*, *I do not know*
- Question to characterize self-reported tendency to take measures to overcome self-control problems (general): "Generally speaking, to avoid putting things off, I often take measures that prevent me from doing so." Subjects then select one of the following items: *Absolutely not like me*, *Very little like me*, *Not really like me*, *A little like me*, *Very similar to me*, *Absolutely like me*, *I do not know*
- Self-reported enjoyment of the task: "Did you enjoy transcribing images of blurry Greek letters?" (0 to 10 scale)
- Instruction recall: "The likelihood of receiving the bonus in Part 1 of the survey was independent of your performance if you were in the" *(a) Performance-Environemnt, (b) Chance-Environment, (c) None of the above*

D Attention Checks

The experiment administered two attention checks.

- **Attention Check 1:** In order to faciliate our research, we are interested in knowing certain factors about you. We are interested in whether you actually took the time to read the instructions; if not, then the data we collect based on your answers will be invalid. So, in order to demonstrate that you have read the instructions, please *ignore* the following question, and then write "I read the instructions" in the box labeled "other". Thank you very much.

  The question read "What do you think is the right interval?" and subjects could choose between multiple intervals or (as instructed) enter something into the "other" box.

- **Attention Check 2:** Please write "the survey ended" into the box on the right:

  Subjects could then either click on one of 10 buttons or, as instructed, write the text into the box.

E Pre-analysis plan

Pre-Analysis Plan

Main Research Questions

This research project aims to answer the following research questions:

1. Do individuals distort beliefs towards believing that their effort is rewarded because they want to enhance their motivation to exert effort in the future (provision of a motive)? Are individuals that receive a signal that indicates that their effort is not rewarded more likely to distort beliefs than individuals that receives a signal that indicates that their effort is rewarded?
2. Are individuals that demand a commitment device more likely to distort beliefs? Are individuals that are bad at the task or dislike the task more likely to distort beliefs?
3. Does motivated belief distortion affect the willingness to redistribute income between two other individuals? Is this effect on redistributive behavior explained through the effect of the motive provision on the belief of being in a state of the world that rewards effort?

Experimental Design

The experiment is separated into four parts. Each part is defined by an activity that may have an effect on either the subjects' own payoff or the payoff of somebody else. At the end of the experiment, one part is randomly drawn on the individual level to become payoff relevant.

Part 1 Subjects are informed that they will perform a real effort task where they transcribe a series of images containing 11 blurry Greek letters for 5 minutes. They are informed that they are matched with one randomly drawn subject (the competitor) that previously participated in the same study (they participated in a pilot study with 102 individuals). The way they are rewarded for this task is uncertain. With 50% chance the computer randomly assigned them either to the *Performance-Environment* or to the *Chance-Environment.* If they are in the *Performance-Environment,* they win a bonus of €7.50 with 80% (20%) chance if they transcribed more (fewer) images than their competitor. If they are in the *Chance-Environment*, they win the same bonus with 50% probability no matter whether they performed better or worse than their competitor. This uncertainty is not resolved. After a comprehension test, subjects carry on with the task.

Part 2 If subjects are in the MOTIVE-GROUP (main treatment variation, 50% chance of being assigned) they are informed that they may redo the *same* transcription task in a later part of the survey. If they were previously assigned to the *Performance-Environment,* they are rewarded with a €5 bonus in the next task

if they transcribe more than 20 images in at most 10 minutes. If they are in the *Chance-Environment* they always get the additional bonus. Note that the environment agents face is the same as in Part 1. Subjects in the NO-MOTIVE-GROUP receive the identical information right before Part 4.

All subjects then receive two pieces of information: (1) whether they performed better than their competitor in Part 1 and (2) whether they won the bonus in Part 1 or not. The signal subjects get is my second treatment variation. They are subsequently asked to state their probabilistic belief (in %) about being in the *Performance-Environment.* Beliefs are incentivized using Karni (2009) with a possibility to win €2.50 and subjects can answer in intervals of 10.

Part 3 All subjects make a redistributive decision. They are informed that they are paired with two subjects that previously participated in the same effort task as in Part 1. One of them received a bonus of €4, while the other one did not receive a bonus. The two individuals that they are paired with are in the same environment as themselves and the same rule (conditional on the environment) as in Part 1 applied to the two individuals. They are then asked whether they want to redistribute the bonus between the two individuals.

Part 4 NO-MOTIVE-GROUP subjects are informed about the following task. All subjects then engage in doing the task.

Questionnaire All subjects finish the experiment with a questionnaire. The questionnaire first asks them to state how many images they plan to transcribe if they were re-invited to redo the task under a piece-rate scheme where they earn €0.25 per correctly transcribed image. It then asks them to specify a "minimum production level" where they will only be paid if they transcribed more images than specified by their "minimum production level". They will actually be reinvited with 5% probability. Before the study concludes with a socio-demographic questionnaire, subjects are asked to recall their signal from Part 2.

Relation between Research Questions and Design

The main treatment variation is whether a subject knows about the future task in Part 4 at the point of belief elicitation in Part 2. This provides the subject with the motive to distort beliefs in order to motivate future effort. Within treatment groups I exploit variation in the signal subjects observe. This allows me to test whether the effect differs by the type of signal subjects receive. Question 2 can be answered by testing for heterogeneity in the main treatment effect based on their propensity to choose a commitment device (Questionnaire), their performance in Part 1, or their self-declared taste for the task (Questionnaire). Finally, the third question is answered by comparing redistribution behavior in Part 3 across treatments.

Sample

500 subjects are recruited via the online platform Prolific for the main experiment. 250 subjects are recruited to serve as recipients of the dictator game. Subjects of the main study receive €3 if they completed the experiment. A subject completed the experiment if they correctly answered two attention checks and if they completed a comprehension test at the beginning of the experiment in at most two tries. Subjects that are not correctly answering all questions after their second try are *not* allowed to complete the study. They receive €0.5. Subjects that are exceptionally fast (completed the study 3 standard deviations below the mean) are also excluded.

Preliminaries

- Main outcomes of interest are (1) the beliefs elicited in Part 2 and (2) the amount redistributed in Part 3
- Main explanatory variables are (1) the treatment group (MOTIVE-GROUP is abbreviated with MG and NO-MOTIVE-GROUP is abbreviated with NG) and (2) whether the subject received a positive (S+) or a negative (S-) signal. A subject received a positive signal if she performed better than her competitor and won the bonus or transcribed fewer images and lost the bonus. A subject received a negative signal if she performed better than her competitor and lost the bonus or transcribed more images and won the bonus.
- Other explanatory variables that are used for the heterogeneity analysis include (1) whether the subject is more likely to self-identify as a meritocrat (dichotomized); (2) Part 1 performance decile of the subject; (3) whether she liked the task (dichotomized); (4) whether she commits herself to transcribe more than half of the images she plans to transcribe if she is reinvited.
- Variables used as control variables are (1) individual characteristics (age, gender, education, nationality, political orientation, risk aversion (Dohmen et al. (2011)), patience (Vischer et al. (2013)); (2) Part 1 performance decile of the subject; (3) winning the bonus; (4) performing better or worse than the competitor (5) false memory (dummy that a subject does not correctly recall her signal)

Main Analysis

The main analysis of the paper is preceded by a descriptive section that plots histograms of beliefs stratified by the signal subjects receive.

The following part lists the analysis I want to undertake to answer the main research questions. Regressions will always use robust standard errors. Each bullet point commences with the outcome and treatment variable in bold.

- **Belief / MG** I first compare beliefs across MG and NG using a two-sided t-test. I then run a two-sided t-test on S+ and a one-sided t-test (H0: Belief | MG ≤ Belief | NG) on S- subjects. I replicate these tests using regression analysis, where I regress elicited beliefs on MG and individual characteristics. Again, this regression is run separately for S- and S+.
- **Logit-Belief/MG,S+** I next compare whether subjects are more likely to distort beliefs relative to a Bayesian benchmark if they receive a negative rather than a positive signal. I regress logit-beliefs on the log odds-ratio of receiving a signal in the *Performance-* relative to the *Chance-Environment.* This variable is interacted with S+ and S-. I *additionally* interact the previously mentioned variables (log odds interacted with S+ and S-) with MG. The regression does not include an intercept. I test the H0 that the interaction with S+ and S- is equal to 1. I test the H0 that each interaction with MG is equal to 0. This coefficient informs how updating behavior differs across MG and NG subjects that received the same signal. Further, I test (t-test) the hypothesis that the magnitude of the coefficient that interacts the log odds ratio of a negative signal with MG is larger than the coefficient that interacts the log odds ratio of a positive signal with MG. The null hypothesis states the opposite.
- **Belief/MG** To test who distorts beliefs, I regress the belief on the treatment variable and on the interaction of the treatment variable with Part 1 performance decile, whether they liked the task and whether the subjects committed herself to work on the future task. I run this regression separately for subjects that received a positive or negative signal.
- **Redistribution/MG** To test whether subjects redistribute less if they are in MG, I test (t-test) whether the amount given to Worker B is different across MG and NG. I do this test separately for S+ and S-. To test whether the treatment effect differs across S+ and S- I run a regression where I interact the MG dummy with the S+ dummy. To assess the role of beliefs in explaining this variation, I conduct a mediation analysis. To that end, I regress redistribution on MG controlling for beliefs to compare this coefficient to the coefficient of a regression that omits the belief variable (inference is made using bootstrapped confidence-intervals). At last, I will interact MG with the variable that indicates whether a subject is a meritocrat. The last two analyses will also be run separately for S+ and S-.

Exploratory Analysis

- **Memory** T-test that checks whether MG subjects are more likely to forget the signal they received. The test is done separately for S+ and S-.
- **Belief / Winning and Rank** I will ask whether there is an effect of winning the bonus on beliefs. I regress beliefs on a dummy that indicates winning controlling for effort decile fixed effects. I will run the same regression with rank as the main explanatory variable. The regression is replicated controlling for the informational content (S+) and including rank as well as winning as explanatory variables. Additionally I run the same regression interacting the explanatory variable of interest (rank or winning) with MG.
- **Redistribution / Winning and Rank** I will ask whether there is an effect of winning on distributive behavior. I regress redistribution on a dummy that indicates winning, controlling for effort decile fixed effects. I will run the same regression with rank as an explanatory variable. I will again perform mediation analysis to test whether the effects are explained by variation in beliefs.
- **Redistribution / heterogeneity** In case there is heterogeneity in belief distortion (see bullet point 3 in the last section), I will test for similar heterogeneity in the treatment effect on redistributive behavior.
- **Beliefs and redistributive behavior** To test for the role of beliefs in redistributive behavior I regress redistribution on beliefs, instrumenting beliefs by S+

Robustness

To assess the robustness of my results, I replicate all regressions using the full set of controls (gender, political, orientation, age, level of education, country of residence, risk aversion and impatience). Furthermore, I will replicate all regressions where I control for Part 1 performance that uses performance decile fixed effects by controlling for Part 1 performance linearly.

Furthermore, I will test to what extent the signal, rank, and winning the bonus can be seen as exogenous. To that end I will run regressions where S+, rank, and winning the bonus is an outcome variable and the above-mentioned individual controls are the main explanatory variable. Any significance should disappear once controlling for performance either by decile fixed effects or linearly. Note that each regression has two explanatory variables (the characteristic and Part 1 performance)

At last I repeat the main analysis with a sample that answered correctly to a basic question about the treatment variation in the last questionnaire.

F Instructions

Figure B1: Screen 1 (all)

Introduction

You are invited to participate in this online study. This is a project conducted by Max Lobeck, a PhD student at Paris School of Economics.
The results of this study will be used in a research project. It is therefore important that you **carefully read and follow the instructions**.

If you choose to participate in this study, you will be asked to do a simple task and answer questions about your opinions on social and economic issues. **The study should take around 30 minutes.**

In this study you have the chance to earn a bonus **in addition to the €3.00 (£ 2.7) you receive for completing** this study. **This bonus can be as high as €7.50.**

The survey is divided into parts. **At the end of the survey, *one* of these parts will be *randomly* selected by the computer.** Each part is equally likely to be selected.

The earnings you made in this randomly selected part will be paid out at the end of the survey. This means that each part is equally likely to determine how much you will earn in this study. It also means **that the amount you earn depends *only* on your decision in this randomly selected part of the survey.**

This study does not use deception. Everything we say in the instructions is implemented in that way. The study got ethical approval by the institutional review board of the Paris School of Economics (Number 2020-015).

The data gathered in this study is subject to EU GDPR regulation. It will only be used for research purposes.

All payments in the instructions are in Euro (€), the final amount will be paid out in British Pounds (£).
We use the following exchange rate: €1 = £0.9.

If you have any questions or concerns, please contact Max Lobeck at max.lobeck@psemail.eu.

If you wish to participate in this study, please consent that you have read the above information and that you agree with participating in this study. Please state your Prolific ID so we can process your payment.
On the last page, you will be given the completion code.

My Prolific ID is:

Consent:

○ I have read the above instructions and would like to participate in the study

Next

Figure B2: Screen 2 (all)

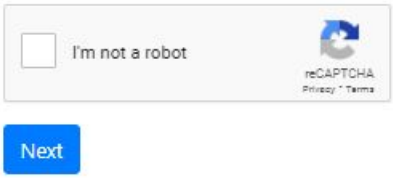


Figure B3: Screen 3 (all)

Part 1 | What is the Task about?

The screen below describes the task you will perform during this part.

After you became familiar with the task, we will inform you how you are rewarded for the task.

Content of the Task

In this task you will be presented with a series of images that contain 11 characters. The characters are either blurry Greek letters or points.

Your job is to transcribe as many of these images as possible in 5 minutes by clicking on the corresponding Greek letters in the bottom row.

This an example of a successful transcription task

Time left to transcribe as many images as possible: **4:43**

Image:

Transcription:

α γ ϵ ϵ χ ϕ α η γ . γ

Recall that you have to transcribe every letter and every point. If you made a mistake you can correct it by clicking on the backwards pointing arrow.

Successfully transcribed images: 0

To submit your transcription and continue with the next image. Please click on the button "Next image" below.

Next image

Even though the letters are blurry, they are depicted in the same way in all images you will face in the course of the task.

If you do not want to transcribe an image, you can always skip that image and continue with the **Next Image** button.

We will now familiarize you with the task by showing you two more examples and giving you the opportunity to transcribe them yourself. *Please note that your performance during this practice phase will not affect any of your earnings.*

Please proceed with the instructions by clicking on the "Next" button.

Next

Figure B4: Screen 4 (all)

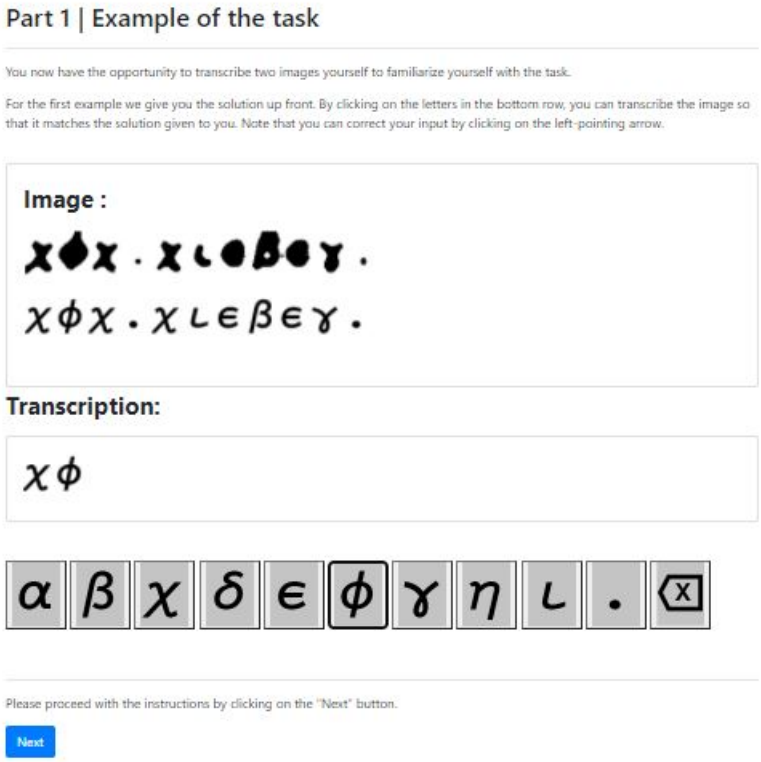


Figure B5: Screen 5 (all)

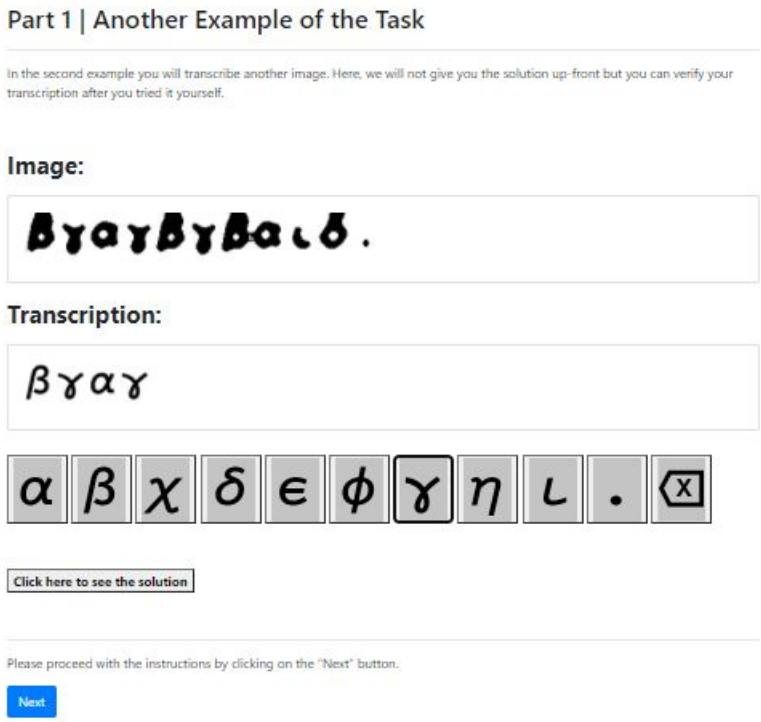


Figure B6: Screen 6 - first part (all)

Part 1 | How are you rewarded for the Task?

Competitor

After you completed the task, the computer will randomly choose your **competitor**.

This competitor is chosen from a pool of participants that just participated in this study on Prolific. These participants completed the same task for 5 minutes after receiving the same information about the task that you just received.

Your performance will be compared against the performance of your competitor. In case you transcribed the same number of images, you rank higher than your competitor if you completed the last correct transcription earlier.

Environment

Before you start the task, the computer will randomly decide in which *Environment* you are performing the task:

- ***Performance-Environment***
- ***Chance-Environment***

The probability that the computer draws either *Environment* is equal to 50%. This means that it is equally likely that you are in the *Chance-Environment* or the *Performance-Environment*.

Thanks to your participation in the task you will have the chance to win a **bonus of €7.50**. The way you win depends on the environment in which you perform the task:

Performance-Environment

- The probability that you win **the bonus *increases*** with the number of images you transcribed correctly
- At the end of the task, the number of images that you transcribed will be compared **to the number of images your competitor transcribed correctly**. The probability that you win the bonus is:
  - Equal to **80%** (8 out of 10) if you transcribed **more images** than your competitor
  - Equal to **20%** (2 out of 10) if you transcribed **fewer images** than your competitor

Figure B7: Screen 6 - second part (all)

Chance-Environment

- The probability that you win **the bonus *is independent*** in the number of images you transcribed correctly
- The probability that you win the bonus is ***always* 50%** (5 out of 10)

After you completed the task, we will inform you whether you won the bonus and whether you correctly transcribed more or less images than your competitor.

Note that this information is informative about the *Environment* in which you actually completed the task.
For example: It is more likely that you completed the task in the *Performance-* rather than the *Chance-Environment* if you did not win the bonus and you transcribed fewer images than your competitor.

Click here to continue with the comprehension questions

Figure B8: Screen 6 - third part (all)

Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.
Note that you will not be able to proceed with the survey if you fail to answer all questions correctly after your *second* try.

Question 1

What is the initial probability in percent (%) that the computer randomly assigns you to the *Performance-Environment*?

Question 2

Suppose that you are in the *Chance-Environment* and you transcribed *more* images than your competitor:
Is the probability that you win the bonus *higher*, *lower* or *the same* than if you had transcribed *fewer* images than your competitor?

- ○ The probability to win is higher
- ○ The probability to win is lower
- ○ The probability to win is the same

Question 3

Suppose that you are in the *Chance-Environment*:
What is the probability in percent (%) that you win the bonus no matter how you performed relative to your competitor?

Figure B9: Screen 8 (all)

Part 1 | Getting Ready

Now that you are familiar with the task and the payment rule, you will proceed with the actual task.

Remember you have 5 minutes to transcribe as many images as possible. On the top of your screen you will see how much time you have left.

Please click on the "Start" button to start the task.

Start

Figure B10: Screen 9 (all)

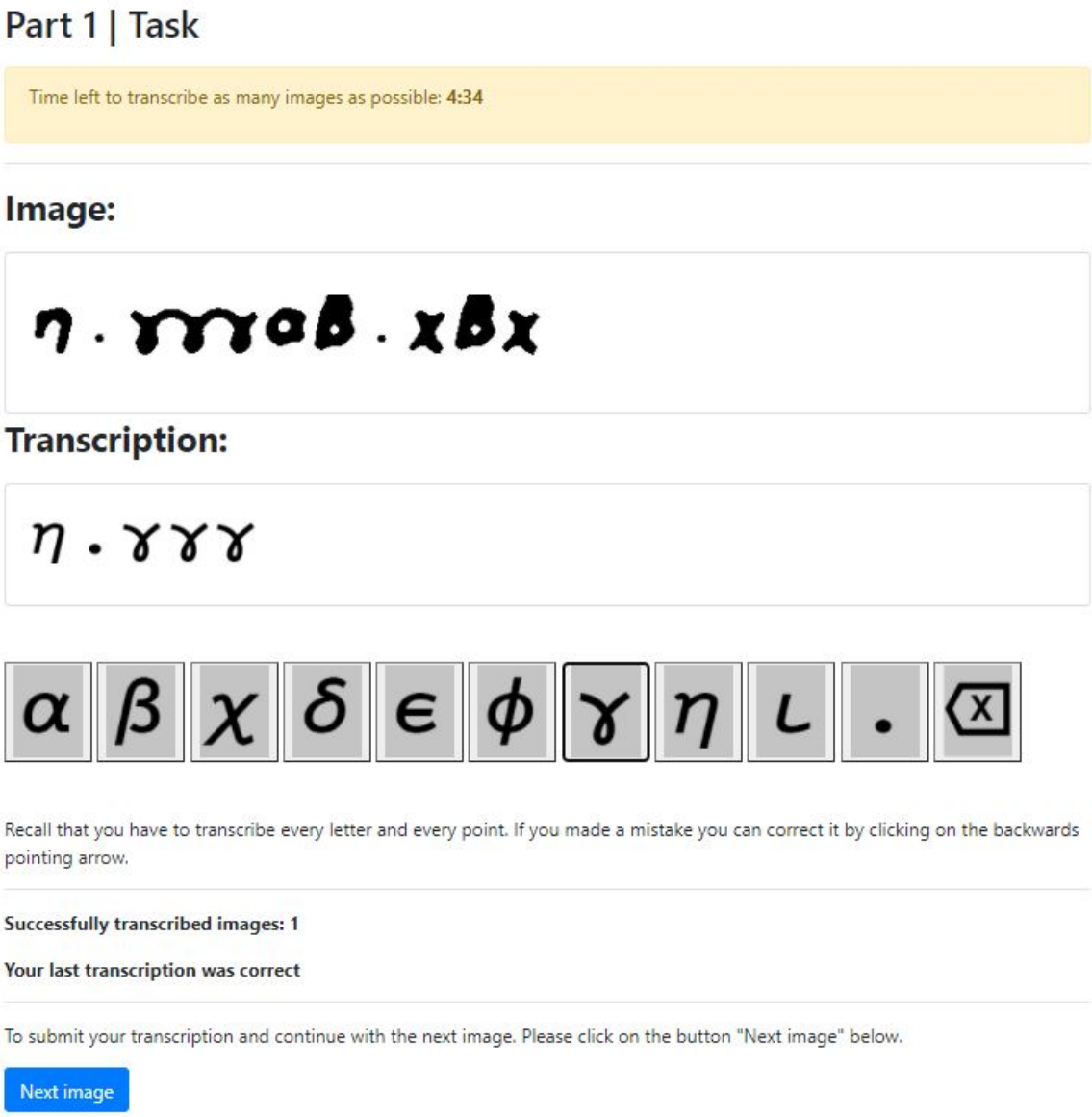


Figure B11: Screen 10 (all)

Part 1 | Task Summary

Your performance in the task

Problem #	Your entry	Solution	Entry was correct
1	ieifihcgcb.	ieifihcgcb.	Yes
2	h.gggab.cbc	h.gggab.cbc	Yes
3	bfcb.cb.chh	bfcb.cb.chh	Yes
4	NA	giaeiaicec.	No

You have transcribed 3 problem(s) correctly.

Please click on the "Next" button to continue with the survey.

Next

Figure B12: Screen 11 (all)

Part 1 | The End

You completed Part 1 of the survey. Please click on the "Next" button to proceed with the second part of the survey.

Next

Figure B13: Screen 21 (motive-group only)

Part 4 | Task

In this part, you will participate to the **same transcription task** as in Part 1. **The environment in which you will perform this task is the one that was drawn at the beginning of Part 1.**

If you performed the task in the *Chance-Environment* in Part 1, then you will perform the next task in the *Chance-Environment*. If you performed the task in the *Performance-Environment* in Part 1, then you will perform the next task in the *Performance-Environment*.

The maximum duration of this part will be 10 minutes. Thanks to your participation in the task you will have the chance to win a **bonus of €5**. The way you win the bonus depends on the environment in which you perform the task:

Chance-Environment

- You ***always*** win the bonus, no matter how many images you transcribed
- Your performance in the task does *not* affect whether you get the bonus or not

Performance-Environment

- You **win the bonus** if you transcribe at least **20 images**
- You **do not win the bonus** if you transcribe fewer than **20 images**
- Your performance in the task does affect whether you get the bonus or not

Figure B14: Screen 22 (motive-group only)

Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

Question 1

What is the activity in the next task?

- O Adding up numbers
- O Solving CAPTCHAs
- O Transcribing images of Greek letters
- O Counting zeros

Question 2

Is the following statement true or false?
Suppose you are in the *Chance-Environment*:
You *always* win the second bonus even if you did not transcribe any images in the next task.

- O True, even if I do not transcribe any images I win the bonus if I am in the Chance-Environment
- O False, I have to transcribe 20 images to win the bonus if I am in the Chance-Environment

Question 3

Suppose you are in the *Performance-Environment*:
How many images do you have to transcribe to win the bonus?

Question 4

Assume that the *Performance-Environment* was drawn at the beginning of Part 1: What bonus-payment rule would apply to you for the next task?

- O Performance-Environment's bonus payment rule
- O Chance-Environment's bonus payment rule
- O Neither

To submit your answers, please click on the "Next" button. The next page will inform you about the results of Task #1.

Next

Figure B15: Screen 12 (all)

Part 2 | Results of the Task

This page informs you about the outcome of the Task you previously completed. Below you find whether you performed better or worse than your competitor and whether you won the bonus.

You *won* the bonus

You transcribed *fewer* images than your competitor

Reminder:

The first line indicates whether you won the bonus or not. If you performed the task in the *Performance-Environment*, the probability that you win the bonus is equal to 20% (2 out of 10) if you transcribed *fewer* images. If you performed the task in the *Chance-Environment*, the probability that you win the bonus is *always* equal to 50%.

The second line indicates whether you transcribed fewer or more images than your competitor.

To continue with the survey, please click on the "Next" button.

Next

Figure B16: Screen 13 (all)

Part 2 | Your estimation of the probability that you are in the *Performance-Environment*

At the beginning of the survey, the computer decided in which environment you perform the task.

In the next screen, you will be asked to report your estimation of the probability that you performed the task in the *Performance-Environment*.

You have the chance to win a bonus equal to €2.50 based on your estimation. Your chance to win this bonus is highest if your estimate is as accurate as possible. It is in your best interest to state what you truly believe is the probability that you are in the *Performance-Environment*. If you want to learn more about the payment mechanism you can click here .

You can proceed by clicking on the "Next" button to provide your answer.

Next

Figure B17: Screen 14 (all)

Part 2 | Your estimation

You can now enter your estimate by choosing one of the options below.

I think the probability that I performed the task in the *Performance-Environment* is:

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ◉ 70% ○ 80% ○ 90% ○ 100%

To submit your answer and proceed with the survey, please click on the "Next" button.

Next

Figure B18: Screen 15 (all)

Part 2 | The End

You completed Part 2 of the survey. Please click on the "Next" button to proceed with the third part of the survey.

Next

Figure B19: Screen 16 (all)

Part 3 | Distribution Decision

A few days ago two individuals, subsequently labelled Worker A and Worker B, were recruited via Prolific to conduct the same task as you (transcribing images of blurry Greek letters for 5 minutes).

Worker A received a bonus of €4.00. Worker B did not receive the bonus. Both workers receive a €1.00 completion reward.

Both workers performed the task in the *same* environment (either *Chance* or *Performance*) as yourself. As a reminder, you just stated that you completed the task in the *Performance-Environment* with 70% probability.

The environment workers faced affects how their bonus was initially distributed.

The rule used to initially distribute this bonus is the same as the rule you faced in Part 1. This means that the probability that Worker A wins the bonus is *independent* of his performance if the task is performed in the *Chance-Environment* and that it is *very likely* that Worker A performed better than Worker B if the task was performed in the *Performance-Environment*.

Please note that they are each other's competitors (Worker B is Worker A's competitor and Worker A is Worker B's competitor).

Description of your decision

- In the next screen, you will be asked to decide about the final payoff of Workers A and B
- The workers were not told who won the bonus nor their environment.
- The workers know that a third person may redistribute this initial allocation of earnings. Your decision is completely anonymous.
- The workers will receive the payment that you choose on the next screen, but they will not receive any further information.

Figure B20: Screen 17 (all)

Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

Question 1

Suppose you performed the task of Part 1 in the *Performance-Environment*:
What is the environment of the two workers?

- ○ Performance-Environment
- ○ Chance-Environment
- ○ Neither

Question 2

Is the following statement True or False?
Suppose you performed the task of Part 1 in the *Performance-Environment*: In that case it is very likely that Worker A transcribed more images than Worker B

- ○ True, it is very likely that Worker A transcribed more images than Worker B
- ○ False, it is equally likely that Worker A transcribed more or less images than Worker B

Please submit your answers by clicking on the "Next" button. Note, that you will not be able to continue with this survey if you did not all answer the questions correctly after two attempts.

Next

Figure B21: Screen 18 (all)

Part 3 | Distribution Decision

How do you want to distribute the bonus?

○ **Worker A** is paid **€4.00** and **Worker B** is paid **€0.00**

○ **Worker A** is paid **€3.00** and **Worker B** is paid **€1.00**

○ **Worker A** is paid **€2.00** and **Worker B** is paid **€2.00**

○ **Worker A** is paid **€1.00** and **Worker B** is paid **€3.00**

○ **Worker A** is paid **€0.00** and **Worker B** is paid **€4.00**

To review the instructions click here

To submit your answer and proceed with the survey, please click on the "Next" button.

Next

Figure B22: Screen 19 (all)

Part 3 | Your Opinion

Before you move on to the next part, we want to ask for your opinion on the following question:

To what degree do you **agree** or **disagree** with the following statement:

If two equally talented workers engage in an *identical* production task, the worker that produced *more units* should be rewarded with a *higher wage* than the worker that produced *fewer units*.

*Using this card, please place yourself on this scale, where **0** means **'totally disagree'** and **10** means **'totally agree'**. You can also use the values in-between to indicate where you fall on the scale.*

totally disagree ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ● 8 ○ 9 ○ 10 totally agree

Please click on the "Next" button to continue with the survey.

Next

Figure B23: Screen 21 (no-motive-group only)

Part 4 | Task

In this part, you will participate to the **same transcription task** as in Part 1. **The environment in which you will perform this task is the one that was drawn at the beginning of Part 1.**

If you performed the task in the *Chance-Environment* in Part 1, then you will perform the next task in the *Chance-Environment*. If you performed the task in the *Performance-Environment* in Part 1, then you will perform the next task in the *Performance-Environment*.

The maximum duration of this part will be 10 minutes. Thanks to your participation in the task you will have the chance to win a **bonus of €5**. The way you win the bonus depends on the environment in which you perform the task:

Chance-Environment

- You ***always*** win the bonus, no matter how many images you transcribed
- Your performance in the task does *not* affect whether you get the bonus or not

Performance-Environment

- You **win the bonus** if you transcribe at least **20 images**
- You **do not win the bonus** if you transcribe fewer than **20 images**
- Your performance in the task does affect whether you get the bonus or not

Figure B24: Screen 22 (no-motive-group only)

Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

Question 1

What is the activity in the next task?

- O Adding up numbers
- O Solving CAPTCHAs
- O Transcribing images of Greek letters
- O Counting zeros

Question 2

Is the following statement true or false?
Suppose you are in the *Chance-Environment*:
You *always* win the second bonus even if you did not transcribe any images in the next task.

- O True, even if I do not transcribe any images I win the bonus if I am in the Chance-Environment
- O False, I have to transcribe 20 images to win the bonus if I am in the Chance-Environment

Question 3

Suppose you are in the *Performance-Environment*:
How many images do you have to transcribe to win the bonus?

Question 4

Assume that the *Performance-Environment* was drawn at the beginning of Part 1: What bonus-payment rule would apply to you for the next task?

- O Performance-Environment's bonus payment rule
- O Chance-Environment's bonus payment rule
- O Neither

To submit your answers, please click on the "Next" button. The next page will inform you about the results of Task #1.

Next

Figure B25: Screen 23 (all)

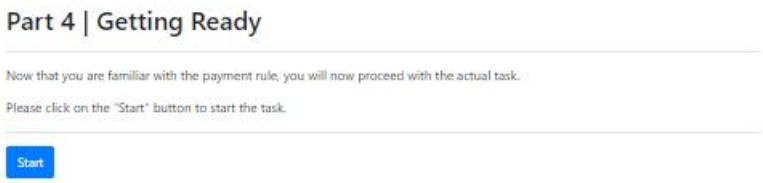


Figure B26: Screen 24 (all)

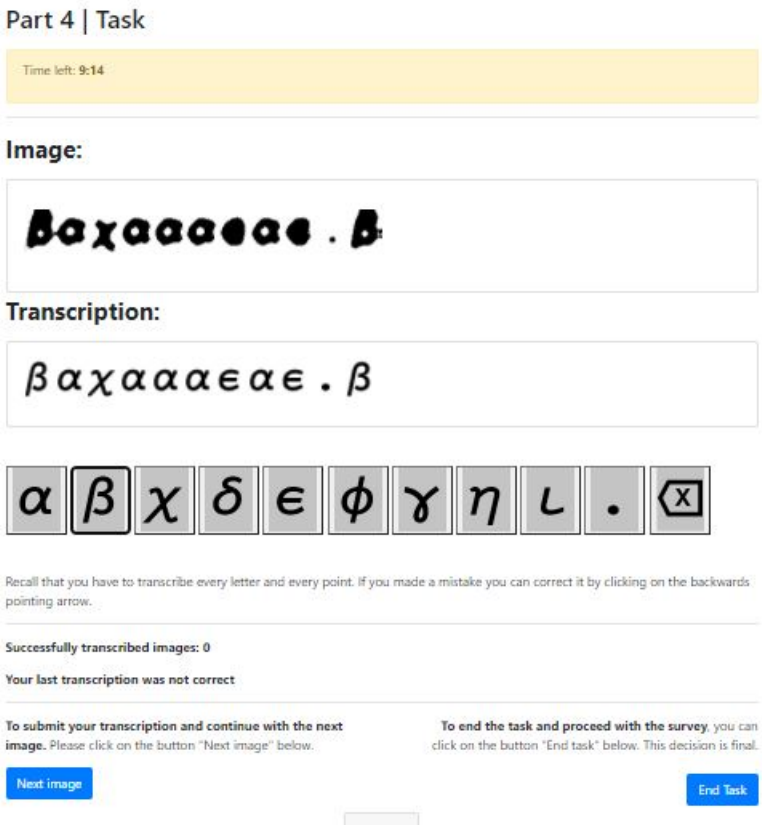


Figure B27: Screen 25 (all)

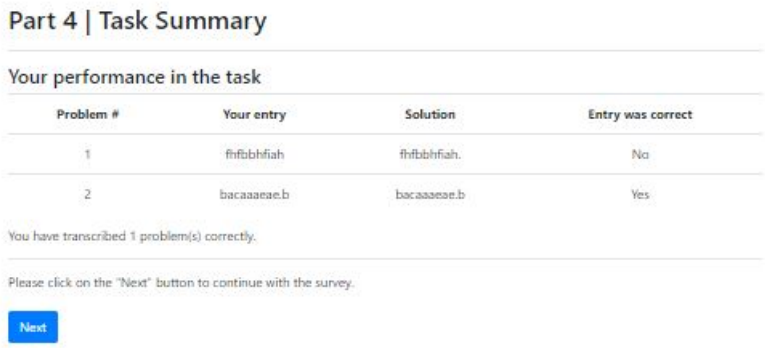


Figure B28: Screen 26 (all)

Part 4 | The End

You completed Part 4 of the survey. Please click on the "Next" button to proceed with the questionnaire part of the survey.

Next

Figure B29: Screen 27 (all)

Questionnaire 1 | Future Task

There is a 5% chance that you will be re-invited to do the same transcription task in 2 weeks.

Specifically, for each image you transcribe you will earn €0.25. This rule will apply to all participants and is independent of the current survey.

In this future task you will be able to transcribe up to 40 images. Hence, you can earn at most €10.00. If you would transcribe 20 images, you will earn €5.00, if you would transcribe 10 images, you will earn €2.50, etc.

In the next screen we will ask you to report how many images you *plan* to transcribe if you were reinvited.

- Note that your answer is not binding
- You will be able to transcribe more or less images in two weeks

By clicking on the "Next" button you can go to the next page and tell us how many images you *plan* to transcribe.

Next

Figure B30: Screen 28 (all)

Questionnaire 1 | Future Task

You can now enter how many images you plan to transcribe if you were reinvited. You can only enter a whole number. The lowest number is 0 (I do not plan to transcribe any images if I were reinvited). The highest possible number is 40.

I plan to transcribe 21 **images**

To submit, please click on the "Next" button. To go back to the instructions, click on the "Back" button.

Back Next

Figure B31: Screen 29 (all)

Questionnaire 1 | Minimum Production Level

In the next screen we will ask you to **define a *minimum production level*** that applies to the task if you are reinvited.

- If your performance in the future task is **lower** than the minimum production level, your earnings will be **zero**
- If your performance in the future task is **higher** than the minimum production level, you will earn **€0.25** per image you correctly transcribed

Example: Assume you define a minumum production level of 7. If you do the task in two weeks and you transcribe:

- 4 images, then your performance is *lower* than the minimum production level, and **you earn €0.00**
- 8 images, then your performance is *higher* than the minimum production level, and **you earn 8*€0.25 = €2.00**
- 12 images, then your performance is *higher* than the minimum production level, and **you earn 12*€0.25 =€3.00**

By clicking on the "Next" button you can go to the next page.

Next

Figure B32: Screen 30 (all)

Questionnaire 1 | Minimum Production Level

You can now enter your *minimum production level*. You can only enter a whole number. The lowest number is 0 (no minimum production level). The highest possible number is 40.

I choose a minimum production level of:

4 images

To submit your *minimum production level* click on the "Next" button. To go back to the instructions, click on the "Back" button.

Back

Next

Figure B33: Screen 31 (all)

Questionnaire 2 | Part 1 Outcome

In Part 1, we asked you to do the transcription task for 5 minutes. After the task you received two pieces of information: (1) whether you transcribed more images than your competitor and (2) whether you won a bonus or not.

Please repeat this information to us.

Did you transcribe more or fewer images than your competitor?

- ● **More** (My performance was higher)
- ○ **Fewer** (My performance was lower)
- ○ **I do not recall**

Please indicate whether you won the bonus for the task of Part 1 or not

- ○ **Yes,** I won the bonus after Task #1
- ● **No,** I did not win the the bonus after Task #1
- ○ **I do not recall**

Please submit your answer by clicking on the "Next" button to proceed with the survey.

Next

Figure B34: Screen 32 (all)

Questionnaire 3

In order to faciliate our research, we are interested in knowing certain factors about you. We are interested in whether you actually took the time to read the instructions; if not, then the data we collect based on your answers will be invalid.

So, in order to demonstrate that you have read the instructions, please *ignore* the following question, and then write "*I read the instructions*" in the box labeled "other". Thank you very much.

What do you think is the right interval?

- ○ 0-20
- ○ 21-40
- ○ 41-60
- ○ 61-80
- ○ 81-100
- Other: read the instructions

Please submit your answer by clicking on the "Next" button to proceed with the survey.

Next

Figure B35: Screen 33 (all)

Figure B36: Screen 34 (all)

Figure B37: Screen 35 (all)

Final Page

You have finished the survey. Thank you for your participation.

This survey has four parts. As mentioned at the start of the survey, we will draw one of these parts randomly and this decision will determine the bonus that is actually paid out.

Below we list the outcome for each of the four parts:

- You got the bonus for the task in Part 1
- You did not win the bonus for your estimation of the likelihood in Part 2
- You decided that Worker A receives €1 and Worker B receives €3 in Part 3
- You got the bonus for the task in Part 4

The computer randomly selected the bonus for the estimation of the likelihood in Part 2 to be the part that becomes payoff relevant.

The amount of €0.0 will be transferred to you in addition to the fixed completion fee of €3.00 in the coming days.

The computer randomly decided that you will not be reinvited to do the task within the next two weeks.

The prolific completion link is: https://app.prolific.co/submissions/complete?cc=8C8F626A

If you have any comments, feel free to contact Max Lobeck at max.lobeck@psemail.eu.