

SPECIAL ISSUE ARTICLE

# Willingness to volunteer among remote workers is insensitive to the team size

Adrian Hillenbrand<sup>1,2</sup>, Tobias Werner<sup>3</sup>  and Fabian Winter<sup>4</sup> 

<sup>1</sup>ZEW – Leibnitz Centre for European Economic Research, L7, 1, 68161 Mannheim, Germany

<sup>2</sup>Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

<sup>3</sup>Max Planck Institute for Human Development, Center for Humans and Machines, Lentzeallee 94, 14195 Berlin, Germany

<sup>4</sup>University of Zurich, Institute for Sociology, Andreasstrasse 15, 8050 Zürich, Switzerland

**Corresponding author:** Tobias Werner; Email: [werner@mpib-berlin.mpg.de](mailto:werner@mpib-berlin.mpg.de)

(Received 31 August 2023; revised 27 June 2024; accepted 22 December 2024)

## Abstract

Volunteering is a widespread allocation mechanism in the workplace. It emerges naturally in software development or the generation of online knowledge platforms. Using a field experiment with more than 2,000 workers, we study the effect of team size on volunteering in an online labor market. In contrast to our theoretical predictions and previous research, we find no effect of team size on volunteering, although workers react to free-riding incentives, and volunteering is perceived as costly. Eliciting workers' beliefs about their co-workers' volunteering reveals conditional volunteering as the primary driver of our results: Workers tend to volunteer more when they believe that others are volunteering, even when doing so is highly inefficient. Using additional experiments, we identify the importance of the task itself as an essential mitigating factor for those results.

**Keywords:** Group size effects; Remote work; Volunteer's Dilemma; volunteering

**JEL Classification:** C72; C93; H41; J4

## 1. Introduction

Volunteering plays a vital role in the organization of many firms. In various situations, tasks and resources are not allocated among employees by some supervisor, but rather employees have to solve the allocation process by themselves. In fact, especially with the rise of remote work, many tasks are organized more informally without a clear hierarchy and require the team members to take their own initiative. While about 4% of the workforce in Europe's biggest economy Germany, worked from home in the years prior to 2020, up to 27% worked remotely in 2021 (Hans Böckler Stiftung, 2021). The US gives a similar picture, with 7.6% of exclusively remote working before the COVID-pandemic and peaking at 31.4% in mid-2020 (Bick et al., 2021). In these situations, volunteering is a natural allocation mechanism that we want to examine in this paper. Completing a task requires time and effort, so each team member would prefer another person to finish it. If the task is completed, the product advances, which yields a higher reputation to the team in the organization and may improve the firm's overall performance in the market. This, in turn, benefits the whole product team as it may improve wages or the job prospects of all team members.

While the described allocation process might seem inefficient due to the challenges of coordination and free-riding incentives, situations like this still occur frequently – and apparently for good reasons – in organizational contexts. For example, many duties in academia are allocated based on voluntary decisions (Babcock et al., 2017), just as the development of open source software projects (Johnson, 2004), the contribution to network technology (Lee et al., 2007), the creation of online knowledge platforms like Wikipedia (Zhang & Zhu, 2011), or modern work allocation mechanisms like the so-called agile project methods, which are commonly used in software development (Hoda et al., 2018). The substantial increase in remote work in recent years might have led to an increase in these allocation mechanisms in the work environment.

The volunteering mechanism does have attractive qualities. Not least, it may reduce the organizational overhead required to organize task allocations. Nevertheless, volunteering in an organizational context is usually not an altruistic act toward others, but often a profit-maximizing individual response to an organizational problem (Murnighan et al., 1993; Kim & Murnighan, 1997). Economic (game) theory thus suggests that the volunteering mechanism also introduces two important obstacles. For one, it creates a coordination problem that the workers must resolve (Erev & Rapoport, 1990). More importantly, however, the individual incentive structure of the firm may give rise to a social dilemma: While working is costly and the exact costs are usually unknown, all team members may enjoy the fruits of labor. This, in turn, leads to the famous Volunteer's Dilemma (Diekmann, 1985).

The strategic analysis of the Volunteer's Dilemma closely resembles several relevant factors for the success of volunteering choices in organizations. First, companies and the teams within the company may be of various sizes, influencing the degree of volunteering (Rapoport, 1985; Otsubo & Rapoport, 2008; Rapoport & Amaldoss, 2013). This has been demonstrated in various laboratory and field experiments with a mostly negative effect of group size on individual volunteering (Latané & Nida, 1981; Diekmann, 1986; Franzen, 1995; Barron & Yechiam, 2002; Przepiorka & Berger, 2016; Goeree et al., 2017; Kopányi-Peuker, 2019; Campos-Mercade, 2021). Second, the individual costs of volunteering might differ for different workers, which also affects individual volunteering choices (Diekmann, 1993; Rapoport & Suleiman, 1993; Kugler et al., 2010; Przepiorka & Berger, 2016). Although laboratory evidence suggests that both factors negatively affect the provision of voluntary work, the prevalence of volunteering in real-world organizations is striking. This begs the question whether economic theory and experimental results from the lab translate into real-world work environments or whether other factors like peer effects, image concerns, or other factors like a different purpose of the task might counteract these problems.

In this paper, we scrutinize these economic arguments against volunteering in the workplace and the existing empirical evidence on volunteering by putting them to a test in real-life workplaces. In a large-scale field experiment with more than 2,000 workers, we analyze the prevalence of volunteering at the workplace and the effect of group size on volunteering behavior. Our main treatment manipulation is thus varying the group size of work teams: we compare the willingness to volunteer for a specific task when working alone to working in small (3 workers), medium (30 workers), and large groups of 300 workers.

In our field experiment, we act as an employer in an online labor market and offer a simple rating task to the online workers. Workers are assigned to a team of a certain group size, and after finishing the individual task, we offer each participant the opportunity to continue working on the task. If at least one person in a group volunteers for the additional team task, each team member receives an additional bonus payment, which resembles the Volunteer's Dilemma. Before the beginning of the team task, participants receive information about their costs and the costs of others as we inform our workers about the time it has taken them to perform the individual task and how long it has taken others. Finally, we elicit workers' beliefs about whether they think another co-worker will volunteer.

Our experimental setting allows us to study the causal effect of differences in team size. By creating a natural yet anonymous work environment, we can exclude reputational concerns as well as personal

relationships between workers, which would impede the analysis of volunteering at more traditional workplaces. Compared to, say, a brick and mortar company, the online labor market provides a unique ecosystem with much tighter control over the task and environment while allowing us to measure individual opportunity costs, beliefs, and other essential variables. At the same time, we are explicitly interested in a natural task environment, and thus did not impose artificial “lab-like” controls on browser use, task completion times, or other things that would be uncommon in this kind of labor markets. Our setting makes individual and team incentives more directly visible compared to a typical work environment. Still, incentives are capturing the trade-off between free-riding and volunteering as in a typical work environment. Workers in our study do not learn whether others have done the task before. While this differs from many classical work environments where the task would be announced as being completed, this design allows us to measure workers’ individual willingness to volunteer, which would not be possible in a work environment where one can only observe whether there was a volunteer or not. We also contribute to the literature on the Volunteer’s Dilemma by providing an application in a real work environment. Compared to existing experiments where effort is purely monetary, we can capture more subtle differences using actual work tasks.

The results of our study stand in stark contrast to the game-theoretical predictions and results from earlier laboratory studies and field settings outside the work context. We find no support for the hypothesis that the group size influences the volunteering decision of our workers. More precisely, in groups of 3, 30, or 300 workers, the volunteering rates range between 51% and 55% and are not statistically different from each other. Also, the costs of volunteering play a minor role, given our proposed cost measure.

We can rule out several possible explanations for this null finding. First, we show that workers are aware of the group size, since around 93% of workers are able to recall the correct group size after the task is completed. In additional treatments in which workers work on the same task alone instead of in a group, we find that (i) workers are less likely to volunteer if volunteering is not compensated, suggesting that the effort is indeed costly, and (ii) workers are more likely to volunteer compared to the main treatments if their payment depends only on their own actions. This does not mean that *every* participant reacted to the strategic environment. Some might have simply enjoyed the task or perceived a duty to help out. It does, however, strongly suggest that our null result is indeed not just a result of strategic confusion or other external factors, but that our participants’ volunteering behavior is indeed insensitive to group size in this workplace environment.

A critical difference between our setting and more typical lab settings is that work has a purpose beyond simply a means to payoffs in an organizational context like ours. In particular, workers in our setting do meaningful work in that their work is used to classify hate speech. In an additional experiment, we shut down this channel by clarifying to workers that there is no further importance to their volunteering decision beyond the monetary incentive. In fact, we observe a significant effect of group size in this setting. It suggests that having a purpose might activate non-strategic motivations to volunteer, making the choice a more individual decision where the costs of the task or norms of behavior might still play a role. However, strategic considerations such as the exact team size are less crucial while in the absence of a purpose more strategic motives are prevalent.

We also observe a form of conditional volunteering. We asked the workers how likely it is that at least one other worker on their team is also a volunteer. We find that those who believe that it is more likely that there is at least one other volunteer are *more* likely to volunteer themselves. Image concerns could explain this seemingly irrational behavior. Subjects who believe that many others probably will volunteer might not want to be perceived (by themselves or us as an employer) as selfish. We show that this form of *conditional volunteering* behavior is a critical driver of volunteering in our work setting. This finding also replicates similar evidence by Rapoport et al. (1989), whose results suggest that participants deviate from equilibrium play in inter-group competitions if they believed that all others contributed as well.

Overall, our results thus suggest that if a company only cares about the task being completed, volunteering can also be justified as an allocation mechanism in large groups, even though theoretical reasoning would predict otherwise. A key difference in our setting compared to the literature on the Volunteer's Dilemma is that the task in itself has a purpose beyond simply ensuring financial gains. According to our results, this does not necessarily lead to a decline in volunteering rates as long as workers have a particular regard for how their work (or not working) is perceived.

The remainder of the paper is structured as follows. In [Section 2](#), we explain the experimental design of the field experiment. Following this, we present our preregistered hypotheses in [Section 3](#). Our hypotheses are empirically tested in [Section 4](#). [Section 5](#) discusses our findings and concludes the paper.

## 2. Experimental design

We study volunteering in a workplace environment through the lens of the Volunteer's Dilemma (Darley & Latané, 1968; Diekmann, 1985). In the Volunteer's Dilemma, a certain number of participants can volunteer to supply a public good to all group members. Volunteering is assumed to be costly. A single volunteer in the group is sufficient to produce a benefit for all its members, which no one receives in case no volunteer can be found. Furthermore, the individual benefit from the public good is greater than the costs of volunteering. This gives rise to the dilemma situation of the game: If there was another volunteer in the game with certainty, the other workers would never volunteer. However, given that the benefit is greater than the costs, workers would prefer to volunteer if all of their colleagues defected. The model captures three essential characteristics of volunteering decisions in the workplace. First, volunteering is chosen simultaneously and without communication, which sets a lower bound for our purposes. Second, there is heterogeneity in, and incomplete information about, the costs of volunteering; and third, firms differ in size. Different tests in the lab (Diekmann, 1986; Franzen, 1995; Goeree et al., 2017; Kopányi-Peuker, 2019) and in field environments (Latané & Nida, 1981; Barron & Yechiam, 2002; Przepiorka & Berger, 2016) have shown that the predictions from the Volunteer's Dilemma are reasonably robust. Thus, one might expect it to be useful in providing predictions in our setting of an online workplace as well.

We use the Volunteer's Dilemma framework to guide our experimental design of the field experiment. In order to establish causal claims, we have to maintain a high degree of experimental control. Online labor markets are therefore not only a convenient, but also a particularly useful environment for our experiment. Importantly, it is a regular and natural workplace for our experimental workers, and workers differ in their effort costs. At the same time, it allows us to set group sizes exogenously. This further differentiates online labor markets from classical work environments, where workers are usually connected through personal relationships and a common history across and within teams. These factors would impede the identification of the causal effects of group sizes, making the use of an online labor market crucial for this study.

In a nutshell, the field experiment consists of two stages (see [Figure 1](#)). In the first part of the job, workers were invited to work individually on a coding task for a fixed payment. Upon joining the job, the workers were randomly matched to one of our treatments in which we varied the group size and the incentive structure.

After completing the first individual task, workers were informed about a second stage. We asked them if they would like to volunteer for a second round of coding, just like the one they had done before, but with a different payment scheme. This second stage implements the actual Volunteer's Dilemma: If at least one worker in the group task volunteered we paid a bonus to all group members. Finally, we elicited the beliefs of the workers about the volunteering decision of their team members, and they had to answer a short questionnaire. All workers received the payoffs a few days after the experiment.

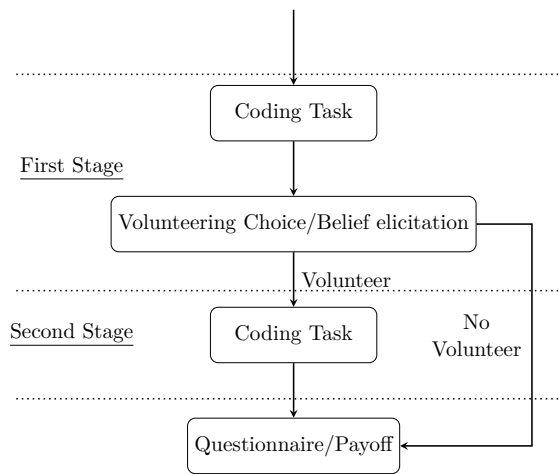


Figure 1. Structure of the experiment

### 2.1. Workers and the online labor market

The field experiment was conducted on *clickworker.de*, an online crowdsourcing marketplace. Crowdsourcing marketplaces allow people to work on tasks that are usually easy to do for humans, but difficult to automate. Most tasks on such platforms require a couple of minutes to complete and include assignments like the processing of images or the cleaning of data (see Difallah et al. (2015) and Jain et al. (2017) for an overview of common tasks). Online labor markets have become increasingly popular in recent years (Difallah et al., 2015), with 0.5% of the US adult population working in the “sharing economy” in 2016 (Farrell & Greig, 2017). For many workers, these jobs serve as a substitute for traditional offline work in times of economic downturn (Borchert et al., 2023). For them, online labor markets are a regular work environment, which makes it a perfect test bed for studying volunteering at the workplace.

Each worker was only allowed to participate once and required to speak German, but we did not impose any further restrictions on the pool of workers. Every active worker on the platform fulfilling those requirements was free to join. Since the platform requires workers to register with a valid German bank account, multiple participations and bots can be practically excluded. In total, 3,344 workers joined the assignment and read the explanation of the task. Overall, 2,203 workers finished the first stage of the field experiment and were then exposed to our volunteering treatments. A drop-out rate of about a third is typical in online labor markets. Altogether, 2,142 workers reached the end of the experiment. For the analysis, we consider only those workers who reached the end of the study. We obtain a diverse sample with a wide range in age, gender, educational status, and employment. Of all participants, 26% report an age between 18 and 25 years, but a sizeable share (14%) is also above 45 years of age. Our sample comprises self-employed, employed, and unemployed workers with various educational backgrounds. Table B.5 in the Appendix reports the sample composition by treatment. It shows that the sample is quite balanced, so different participant characteristics should not drive potential treatment differences.

### 2.2. The advertised job

We offered a standard job to all workers active at the time through an advertisement on the platform. Notably, there was no mention of an experiment or anything similar. The workers were asked to rate user comments from another study, a task that is often found in online labor markets. We provided a short description of the task and informed the workers that they could earn 0.90 € and how long it

would take them to complete it. While this wage may seem low, it was the standard wage set by the platform, and since we wanted the task to feel as natural and standard as possible, we had to stick to the platform's suggestion. As we will discuss later, the workers felt this was a decent enough wage to complete the task.

Once they had clicked on our link on the platform, they were redirected to our oTree server (Chen et al., 2016). Workers then received detailed instructions about the task where we also made them aware that they would be working in a team for this assignment. We clarified that they would first be working alone on the described assignment. After completion they were offered a voluntary group project (see Appendix E for the full instructions).

**The Task** Workers were asked to evaluate user comments from an online forum. Each comment was made with reference to a picture and possibly comments from other users in this online forum. The picture's theme was always related to migration, refugees, or cultural differences. The workers rated the comments regarding the expressed sentiment and evaluated whether the comments contained hate speech. The coding scheme and a screenshot of the task can be found in Appendix E.2. In both possible experiment stages, 30 distinct comments had to be rated, which were randomly drawn from a set of 13,356 comments. All comments had been collected as auxiliary data in a different study and are not part of our research question. Companies and corporations frequently use online labor markets for similar tasks to better understand customer comments or reviews. According to Difallah et al. (2015), those verification and validation tasks belonged to the most common assignments between 2009 and 2014 on Mturk, a US-based competitor of *clickworker.de*. According to *clickworker.de*'s own information, verification tasks and sentiment analyses are the most common tasks in their industry. Thus, we are confident that most workers were familiar with this type of task and perceived it as a regular assignment rather than part of a research project. The ratings of the comments will be used in a series of studies on hate speech for which the original comments had been collected as a dependent variable in the first place (Álvarez-Benjumea & Winter, 2018, 2020; Álvarez-Benjumea, 2020). Thus, the work was indeed meaningful and important.

### 2.3. The volunteering decision

After finishing the first stage of the experiment, we explained to the workers that we needed exactly one volunteer in their team to ensure data quality and to be better able to evaluate the quality of the ratings within their group. We clearly stated that one volunteer within the group was sufficient for the task. Each worker received the offer to volunteer in their team and to continue to work on the task for another round of 30 comments in the second stage. If at least one person in the group volunteered, all members received a bonus payment of 0.90 €. If no person in the group volunteered, no group member received any bonus. The bonus payment did not increase if more than one person volunteered. Furthermore, we explained to the workers that, even if they had not volunteered themselves, they might still receive the bonus payment if one of their teammates volunteered. Importantly, workers were not made aware if others already did the task before making their own choice. This allows us to measure the actual willingness to volunteer. To avoid reputation effects, we clarified that their decision would not have an influence on the payoff of the first stage or their user rating in the online labor market. The volunteering decision in the second stage will be our key dependent variable in the analysis in Section 4. A screenshot and a translation of the decision screen can be found in Appendix E.3.

We used the time spent in the first part of the job as a cost measure for volunteering. In our context, time spent on a job is a fair measure of opportunity costs, because those who spend more time on the task miss out on more opportunities to work, for example, on another job on the platform or on more leisure. As we show later, workers differ substantially in the time it takes them to complete the task. Since both parts of the job consist of the same task, the time spent on the first part is also a



good predictor for time spent on the second part ( $\rho = 0.73, p < .001$ , Spearman's rank correlation coefficient). Workers were made aware of the time it had taken them to complete the first stage.

In accordance with our theoretical model, we induced commonly known beliefs about the distribution of costs of other workers before making the volunteering decision. To this end, we informed them that other workers usually required between 7.5 to 15 minutes to complete the rating of the comments.<sup>1</sup> This gave workers a rough estimate as to whether they had high or low costs of volunteering relative to the other workers.

## 2.4. Treatments

**Volunteer's Dilemma Treatments** Workers in our main treatments faced team incentives in the form of a Volunteer's Dilemma. Within the field experiment, we varied three group sizes, the SMALL GROUP with 3 workers, the MEDIUM GROUP with 30 workers, and the BIG GROUP with 300 workers.<sup>2</sup>

The instructions were adapted accordingly for each group size. Since the information on the size of the group is the key manipulation in our study, we made sure that the information was as clear as possible for our workers. The group size was mentioned several times, in particular, right before workers made a choice.

**Baseline Treatments (N=1)** In order to provide a benchmark for volunteering rates in our setting, and to establish whether volunteering is actually costly in this task, we designed two control treatments: INCENTIVIZED and UNINCENTIVIZED. The basic setup in this study was identical to our main experiment, the difference being that the workers did not face a Volunteer's Dilemma. That is, the instructions were the same as in the Volunteer's Dilemma treatments, including the fact that workers operated in a team. We pointed out, however, that their actions did not influence the payoffs of their team members, and vice versa.

Workers were notified at the end of the first stage that we needed a volunteer to continue working on the task.<sup>3</sup> We varied two conditions. In the INCENTIVIZED condition, subjects were paid the same bonus as in the Volunteer's Dilemma treatments (90 cents) for completing the second part. Yet, if they did not volunteer, they did not receive the bonus even if another team member volunteered. This condition would provide an upper bound of volunteering without any team incentives. In the UNINCENTIVIZED condition, there was no bonus, which allows us to control for intrinsic or non-monetary motivations to finish the task. The volunteering rates of our main treatment, where strategic considerations play a role, should lie between these two conditions.

The assignment was made unavailable once each main treatment had reached 600 workers who had finished the first stage.<sup>4</sup> For the baseline treatments, the task was made unavailable after 200 workers had made a volunteering decision. We remain with the treatment composition shown in Table 1.

## 2.5. Belief elicitation

We elicited the beliefs of the participants about the volunteering decision of other members in their group. We asked each participant how likely it is that at least one of their team members volunteers

<sup>1</sup>These numbers were collected in a pilot study with 100 workers. We showed the workers the 20th and 80th percentile of the time values, but referred to these values as the time it took "most workers"

<sup>2</sup>To implement a truly random and independent draw of the group size, there was an individual random draw for each worker. Then, for each worker a random team of the specific size was generated and the worker's payments were based on the actions of these workers. This method is similar to the method used by Boosey et al. (2017).

<sup>3</sup>Note that in the main treatments, we explicitly told the workers that we only required one person to volunteer. In this baseline case, we did not specify how many volunteers were required, but rather that we needed volunteers. The INCENTIVIZED condition can be understood as a Volunteer's Dilemma with a single worker.

<sup>4</sup>The main treatments were limited to 600 workers who completed the first stage, but due to dropouts in the second part of the experiment, the final count in each treatment fell slightly below 600.

**Table 1.** Number of observations for each treatment

Group Size	Number of Observations
UNINCENTIVIZED ( $N = 1$ )	192
INCENTIVIZED ( $N = 1$ )	196
SMALL GROUP ( $N = 3$ )	588
MEDIUM GROUP ( $N = 30$ )	582
BIG GROUP ( $N = 300$ )	584

Note: The number of workers for each treatment, with  $N$  being the group size.

for the task. Participants reported the probability on a scale from 0% (“For sure no other person”) to 100% (“Surely at least one other person”) using a slider. The slider did not have an initial value to avoid any anchoring effect.<sup>5</sup> The belief elicitation was incentivized using the binarized scoring rule and the possible bonus is 0.90€ (Hossain & Okui, 2013).<sup>6</sup> The instructions were simple and natural. To increase truth-telling, we used the framing by Danz et al. (2022), and told participants that it is in their best interest to report their actual belief and that the chance to receive the bonus depends on the accuracy of their guess. We randomized the order in which we asked participants for their volunteering decision and their belief. Half of the participants were asked just before their volunteering decision about their belief. The other half answered the belief question after their volunteering decision but before the potential second round of coding comments. It allows us to investigate possible order effects. As the order in which we ask participants for their volunteering decision and their belief plays a negligible role in the results, we pool all observations along this treatment variation for the analysis in Section 4.4.<sup>7</sup>

After the participants reported their beliefs, we also elicited how certain they were about this probability. To measure this “cognitive noise,” we followed the approach by Enke & Graeber (2023) and Enke et al. (2023).

## 2.6. Questionnaire

At the end of the field experiment, workers completed a questionnaire about their economic preferences (Falk et al., 2023) and their sociodemographic/economic background. We also asked the participants several questions about the task they had to perform. We use the answers as an additional measure of cost. Furthermore, we asked participants if they could correctly recall the size of their team. We kept the initial questions to a minimum so as not to disturb the impression that the experiment was anything other than a normal job.

## 3. Hypotheses

The incentive structure in our field experiment resembles a Volunteer’s Dilemma (Diekmann, 1985) with incomplete information and heterogeneous costs. For our model, we thus focus on the setup by Weesie 1994. We introduce the model’s key features in the following to derive our hypothesis. The key insights from the model are that volunteering decreases in the group size and in the costs of volunteering. The full model, including all proofs, is provided in Appendix A.<sup>8</sup>

<sup>5</sup>A screenshot is provided in Appendix E.

<sup>6</sup>The winning probability was calculated as  $p = 1 - (V - \frac{\text{Belief}}{100})^2$  where  $V$  is equal to one if there was another volunteer in the team and zero otherwise.

<sup>7</sup>See Appendix C.2 for an analysis of the order effects.

<sup>8</sup>While this might seem intuitive for most readers, it is worth mentioning that this prediction reverses the predictions under common knowledge of costs, where those with higher costs volunteer more in equilibrium.



To be more formal, there are  $N$  players in the game.<sup>9</sup> Each player  $i \in \{2, \dots, N\}$  decides simultaneously to volunteer ( $a_i = V$ ) or to defect ( $a_i = D$ ). If at least one player in the game volunteers, all players receive a benefit of  $b_i$ . Volunteering is costly and the costs are denoted by  $c_i$ . In line with Weesie (1994), we assume that  $\gamma_i := \frac{c_i}{b_i}$  follows some arbitrary probability distribution  $\gamma \sim \mathcal{F}$  with a continuous probability density function  $f$  and that  $b_i > c_i > 0 \forall i$ . The payoff  $\pi_i$  of worker  $i$  when  $X_{-i}$  others volunteer is

$$\pi_i = \begin{cases} 0 & \text{if } a_i = D \text{ and } X_{-i} = 0 \\ b_i & \text{if } a_i = D \text{ and } X_{-i} > 0 \\ b_i - c_i & \text{if } a_i = V. \end{cases}$$

The payoff structure gives rise to the dilemma situation of the game: If there were another volunteer in the game for sure, workers would never volunteer. However, given that the benefit is greater than the costs, workers would prefer to volunteer if all of their colleagues defected. Formally, player  $i$  volunteers if and only if its expected benefit is weakly greater than the benefit from defecting.

$$\begin{aligned} EU(V) &\geq EU(D) \\ \Leftrightarrow b_i - c_i &\geq b_i P(X_{-i} > 0) \\ \Leftrightarrow \gamma_i &\leq 1 - P(X_{-i} > 0) \end{aligned} \quad (1)$$

Thus, she volunteers if and only if the cost-benefit ratio  $\gamma_i = \frac{c_i}{b_i}$  of player  $i$  is weakly below the probability that there is no other volunteer. Weesie (1994) shows that there exists a pure strategy equilibrium where players with low cost-benefit ratios volunteer, while those with a  $\gamma_i$  above some threshold do not volunteer.

In our field experiment, we can elicit the main parameters of the model. The additional bonus of 90 cents is the benefit  $b_i$ . Volunteering for the task costs time and effort. We argue that those costs to volunteer ( $c_i$ ) can be approximated by the time it takes participants to complete the first part of the experiment and additional survey measures that we elicit in a questionnaire. Also, we directly ask participants for their belief about  $P(X_{-i} > 0)$ , that is, the probability that there is at least one other team member volunteer. As a result, our experimental design allows us to test a wide range of hypotheses, which are derived from the equilibrium analysis by Weesie (1994). We provide further details on the theoretical foundations in Appendix A.

In our baseline treatments, volunteering is an individual choice. It allows us to identify if volunteering is indeed costly, and participants react to the dilemma component of the game. If participants perceive the task as costly, they should volunteer less if there is no additional benefit. Thus, we expect volunteering rates in INCENTIVIZED to be higher than in the UNINCENTIVIZED treatment. The group size treatments introduce the Volunteer's Dilemma and are thus our main treatments. Accordingly, we expect that the volunteering rate is lower in the main treatments compared to the INCENTIVIZED treatment, in which participants only influence their own benefit if they volunteer. Yet, as we expect that volunteering is perceived to be costly, we hypothesise that more participants volunteer in the

<sup>9</sup>The game can be easily extended to allow for stochastic group sizes. The details are also provided in the respective section of the Appendix.

main treatments compared to the UNINCENTIVIZED treatment. This argument is summarized by Hypothesis 1.<sup>10</sup>

**Hypothesis 1.** The volunteering rate in the Volunteer's Dilemma treatments is higher than in the UNINCENTIVIZED treatment but lower than in the INCENTIVIZED treatment.

Participants with lower costs should volunteer more for the additional task in the main treatments. This prediction follows intuitively from Equation 1 as the expected utility from volunteering is decreasing in the costs. A formal proof is provided in Appendix A.

**Hypothesis 2.** Workers with lower costs are more likely to volunteer.

From a theoretical perspective, the share of volunteers decreases in the group size. Large groups give rise to a bigger “diffusion of responsibility”, and players volunteer less on an individual level. We expect that the effect carries over to our workplace environment.

**Hypothesis 3.** The volunteering rate decreases in the group size.

We elicit the participants' beliefs that there is another person in the team who volunteers. We can verify whether participants have correct beliefs using the average volunteering rate in each treatment. Furthermore, the expected value from defecting is increasing in this belief (see Equation 1). For higher beliefs, there exists a higher subjective probability that a participant receives the benefit without volunteering herself. It follows that participants with higher beliefs should volunteer less, as they can avoid incurring the volunteering costs while receiving the same benefit.

**Hypothesis 4.** Participants who believe that there is another volunteer are less likely to volunteer themselves.

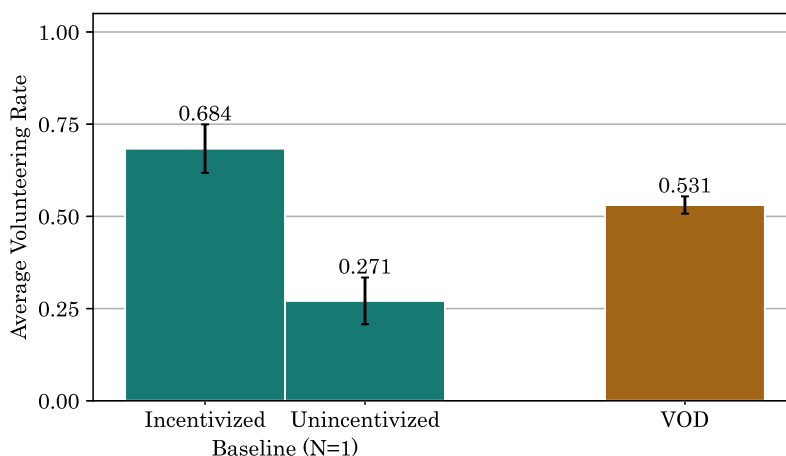
Clearly, all these hypotheses build on purely monetary-driven arguments. In a work environment, multiple additional factors, such as peer effects or image concerns, might play a role. These factors might influence the link between beliefs and actions in different ways.

## 4. Results

We will discuss our empirical results as follows: In the first step, we demonstrate that workers perceive the task as costly and react to free-riding incentives. We then follow the hypotheses outlined in Section 3. To understand the mechanisms underlying our results, we investigate the stated beliefs of subjects about their pivotality to provide the good and the interplay between beliefs and actions. We will finally discuss additional experimental evidence, suggesting that the perception of one's work as “meaningful” may transform the “games” people play in social interaction.<sup>11</sup> Appendix B contains additional detailed empirical analyses supplementing our results.

<sup>10</sup>Further details on the theoretical model are provided in Appendix A. The pre-registered hypotheses can be found under <http://dx.doi.org/10.17605/OSF.IO/7RQVH> and <http://dx.doi.org/10.17605/OSF.IO/8TZ57>. The wording of the hypotheses was adapted, and we added hypotheses about beliefs based on theoretical results to fit the structure of the paper. Note that some results, especially those regarding the case of an uncertain group size, are discussed in a previous retired version of the paper, which we do not further explore in this version <https://osf.io/4k5y6>.

<sup>11</sup>Furthermore, we provide additional robustness tests and replications for our results in Appendix C.



**Figure 2.** Volunteering rates in the incentivized and the unincentivized baseline treatments in comparison to the main Volunteer's Dilemma (VOD) treatments. The error bars represent the 95% confidence intervals

#### 4.1. Workers are sensitive to incentives and strategic situations

To satisfy our first identifying assumption, we have to establish whether the volunteering choice was actually costly. We therefore compare two baseline treatments in which we asked participants whether they wanted to volunteer in a second coding round. In these baseline treatments, the payment only depends on the worker's decision and not on other workers.<sup>12</sup> In the INCENTIVIZED condition, participants were paid an additional 0.90€ in case they volunteered for a second round of coding. In the UNINCENTIVIZED condition, they were only paid for the first round.<sup>13</sup> The volunteering rate in the INCENTIVIZED version was 68.4%, but only 27.1% in the UNINCENTIVIZED treatment (See Figure 2,  $p < 0.01$ ,  $\chi^2$ -test). This allows us to conclude that the task was indeed perceived as a costly effort and provides the basis for the following results.

Our second identifying assumption was that our participants react to the strategic situation of the Volunteer's Dilemma and show different volunteering rates than in the individual choice situation. We expected that volunteering rates in the main treatments fall between those in the INCENTIVIZED and the UNINCENTIVIZED individual choice condition. This is also the case. Pooling the data for all main treatments, the volunteering rate is significantly lower, at 53.1%, than in the INCENTIVIZED condition ( $p < 0.01$ ,  $\chi^2$ -test) and significantly higher than in the UNINCENTIVIZED individual choice condition ( $p < 0.01$ ,  $\chi^2$ -test, see Figure 2). Hence, we find clear support for Hypothesis 1.

**Result 1.** Volunteering is costly, and workers react to free-riding incentives.

#### 4.2. Heterogeneity in costs to volunteer

Hypothesis 2 predicts that participants with higher costs are less likely to volunteer. As explained in Section 2, we argue that individual costs of volunteering can be approximated by the time it takes a participant to finish the first stage of the field experiment. Most participants required between 8.23 (20th percentile) and 16.52 minutes (80th percentile) to complete the first stage, with an average of

<sup>12</sup>Note that we kept the framing of them being part of the team to keep the decision environment constant.

<sup>13</sup>As explained in Section 2 for the main treatments, we only consider those participants who had reached the end of the experiment. Out of 400 participants who started the experiment, we ended up with 388 observations: 196 in the INCENTIVIZED condition and 192 in the UNINCENTIVIZED condition.

**Table 2.** Logistic regression to estimate the volunteering choice as a function of different cost measures

	<i>Dependent variable:</i>		
	Volunteering Choice		
	(1)	(2)	(3)
TIME	−0.007 (0.048)		−0.005 (0.050)
COST INDEX		−0.375*** (0.050)	−0.375*** (0.050)
Constant	0.123*** (0.048)	0.126*** (0.049)	0.126*** (0.049)
Observations	1,754	1,754	1,754
Log Likelihood	−1,212.443	−1,183.000	−1,182.994
Akaike Inf. Crit.	2,428.886	2,370.001	2,371.989

Notes: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ , Standard errors in parentheses.

14.50 minutes. Figure B.1(a) in the Appendix presents the estimated distribution of completion times by treatment.

The data does not support our Hypothesis 2 since the effect of costs operationalized as time is insignificant. Model 1 in Table 2 shows the results of a logistic regression model and estimates the probability of volunteering as a function of the z-standardized time. Importantly, the coefficient of TIME is insignificant.<sup>14</sup>

We constructed a z-standardized additive “COST INDEX” from several control questions in our post-experimental questionnaire as an additional measure of subjective costs. We asked participants on a 7-point Likert scale whether they perceived the task as exhausting ( $\mu = 3.01$ ), interesting ( $\mu = 2.91$ , reverse-coded), or emotionally challenging ( $\mu = 2.30$ ), and whether it was important to them to contribute to “better data quality” ( $\mu = 2.44$ , reverse-coded). The subjective costs have a substantially meaningful and highly statistically significant negative effect on the probability to volunteer (see Model 2 and 3 in Table 2 and Table D.3). The effect of costs on volunteering are similar across treatments and visualized in Figures B.3 and B.4 in the Appendix.

We therefore conclude that

**Result.** The probability to volunteer decreases in the subjective costs of volunteering.

#### 4.3. Insensitivity to team size in the volunteers dilemma

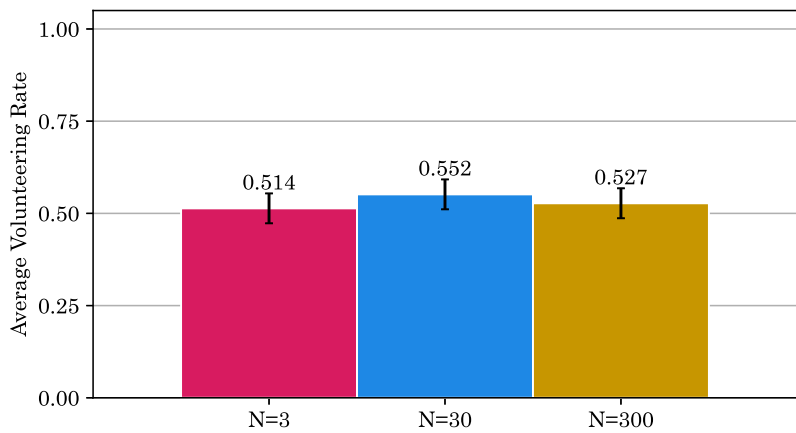
Hypothesis 3 predicts that the volunteering rate decreases in the group size. This is clearly not the case (see Figure 3 and Tables B.1 and B.2). Volunteering rates are relatively high and statistically indistinguishable with regard to group size (all  $p > 0.1$ ,  $\chi^2$ -tests).<sup>15</sup> We therefore conclude that

**Result.** The volunteering rate does not vary in the group size.

This result is surprising, given the extensive literature on the volunteer dilemma. It constitutes our main result and shows that our participants react to economic incentives and the strategic nature of the Volunteer’s Dilemma (Result 1), but that team size seems irrelevant.

<sup>14</sup>This is also the case for the estimated average marginal effect, which can be found in Table D.3 in the Appendix.

<sup>15</sup>We also test for difference in volunteering across group sizes using a logistic regression in Table B.3. The average marginal effect estimates in Table B.4 show that the effect of group size is small and not statistically significant at any conventional level.



**Figure 3.** The average volunteering rate across the different group sizes. The error bars represent 95% confidence intervals

**Table 3.** Belief overview by treatment

Group size	Correct belief	Average belief	Share with correct* belief
SMALL GROUP (N = 3)	76.34	60.66	0.11
MEDIUM GROUP (N = 30)	100.00	70.24	0.30
BIG GROUP (N = 300)	100.00	70.89	0.30

\*The correct belief is calculated based on the average volunteering rate in the respective treatment. Here, we define the belief of a participant as correct if it is in the interval of +/- 5 pp around the correct belief.

4.4. Beliefs

A potential explanation for this surprising result might be that our participants had similar beliefs about their pivotality across treatments. In this section, we show that while beliefs differ across treatments in their level and marginal effect on the likelihood of volunteering, this difference is insufficient to account for the striking similarities in volunteering rates. To do so, we compare the beliefs with the actual probability that at least one other person volunteers for the given group size.<sup>16</sup>

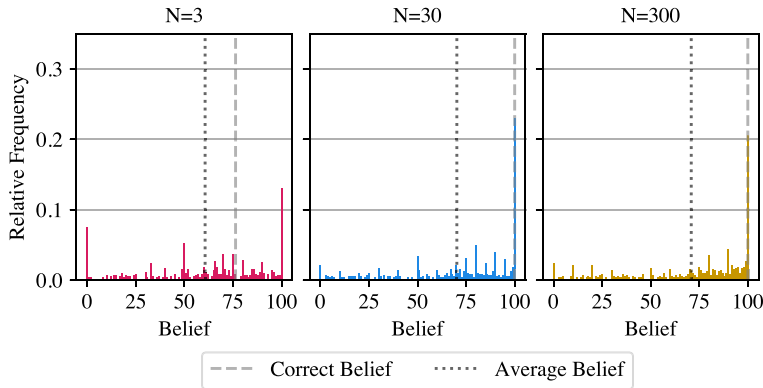
Table 3 reports the results. In each treatment, participants report beliefs that are, on average, smaller than the correct belief. As such, they underestimate the probability that there is at least one other volunteer, thereby overestimating their own true pivotality. The differences are statistically significant ( $p < 0.01$  for all treatments).<sup>17</sup> Hence, participants do not form correct beliefs about the volunteering decisions of other team members.

Figure 4 shows the discrete distribution of the beliefs by group size. There exists a considerable variation across beliefs. While about a third of participants in  $N = 30$  and  $N = 300$  report beliefs close to the correct belief, beliefs are, on average, too low. In  $N = 3$  even fewer participants report a belief close to the true probability that one other team member volunteers. This is also driven by a large share of participants who believe that no one else volunteers.

**Result.** Beliefs about volunteering are on average incorrect.

<sup>16</sup>The probability that one other person volunteers is given by  $P(X_{-i} > 0) = 1 - (1 - s)^{N-1}$  where  $s$  denotes the share of volunteers in a given treatment.

<sup>17</sup>We regress the difference between correct and reported beliefs in a Tobit regression on a constant.



**Figure 4.** The belief that another person in the group volunteers by group size. To calculate the correct belief we use the average volunteering rate in the treatment

Next, we consider differences in beliefs across treatments. Participants in the  $N = 3$  treatment report statistically significantly smaller beliefs than in the other two treatments ( $p < 0.01$  for both comparisons, two-sided Mann–Whitney U test). There are no differences in beliefs between  $N = 30$  and  $N = 300$  ( $p = 0.64$ , two-sided Mann–Whitney U test).

**Result 5.** The belief that someone else volunteers is lower in small teams compared to larger teams.

The differences in beliefs make intuitive sense, given the observed volunteering rates in the field experiment. As volunteering rates are similar, the probability that another person in the team is a volunteer increases in the group size.<sup>18</sup>

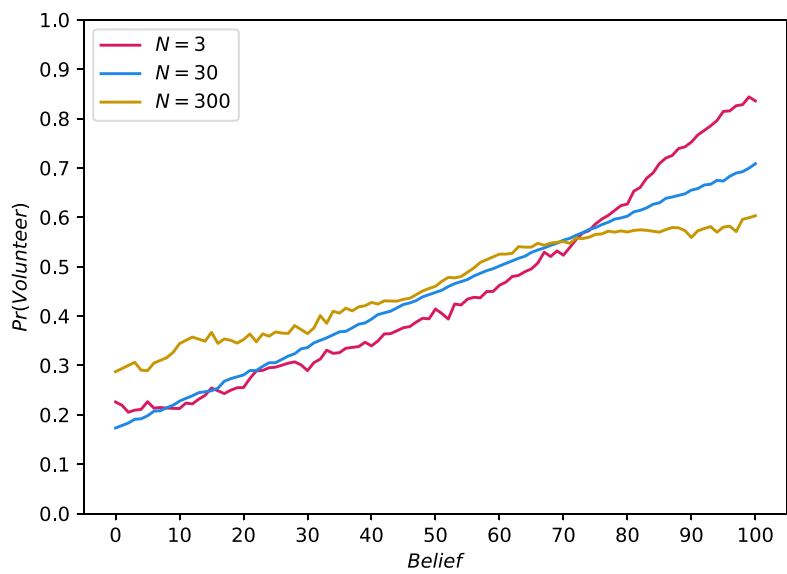
From a participant's perspective, one would expect that lower beliefs lead to higher volunteering and, thus, to see more volunteers in small groups. However, participants volunteer at similar rates across treatments. A possible explanation is that a non-trivial link exists between actions and beliefs in the workplace environment, which is not captured by the classical Volunteer's Dilemma framework. We investigate the relationship between beliefs and actions in the following section.

#### 4.5. Conditional volunteering

We now focus on how beliefs translate into actions. As discussed in Hypothesis 4, differences in beliefs should translate into differences in behavior. The higher the probability that someone else in the team volunteers, the higher the likelihood that the player still receives the benefit without volunteering herself. From a purely monetary perspective, subjects who are certain that another participant volunteers (i.e., report a belief of 100%) should never volunteer themselves as volunteering is costly and the benefit does not increase if there are multiple volunteers. On the other hand, participants with pessimistic beliefs about the volunteering of others should be

<sup>18</sup> Furthermore, we observe a higher degree of cognitive uncertainty about the belief in the SMALL GROUP compared to the other two group sizes ( $p < 0.01$ , two-sided MWU tests). There is no difference between the MEDIUM GROUP and the BIG GROUP ( $p > 0.1$ , two-sided MWU test). Following the argument by Enke & Graeber (2023), it indicates that subjects shrink their beliefs more towards an ignorant prior. While this ignorant prior is unobserved in our experimental design, it is reasonable to assume that it is 50% given the binary outcome. As a result, cognitive uncertainty can also explain the belief differences between the SMALL GROUP and the other group sizes. In line with Enke & Graeber (2023) we find a hump-shaped relationship between reported probabilities and cognitive uncertainty. In other words, participants who report beliefs close to 50% tend to encounter a higher degree of cognitive noise. The effect is especially strong in smaller groups. See Figure B.2 for further details.





**Figure 5.** Probability to volunteer based on belief about at least one other worker volunteering. Fitted values from a local linear regression with Wang Ryzin Kernel and leave-one-out cross-validated bandwidth

**Table 4.** The Volunteering choice explained by Beliefs and all treatment dummies in a linear probability model with robust standard errors

	Dependent variable:		
	Volunteering Choice		
	(1)	(2)	(3)
BELIEF	0.005*** (0.0004)	0.005*** (0.0004)	0.007*** (0.001)
MEDIUM GROUP		−0.009 (0.027)	0.090 (0.062)
BIG GROUP		−0.037 (0.028)	0.205*** (0.064)
MEDIUM GROUP × BELIEF			−0.002** (0.001)
BIG GROUP × BELIEF			−0.004*** (0.001)
Constant	0.203*** (0.026)	0.214*** (0.029)	0.108*** (0.040)
Observations	1,754	1,754	1,754

Notes: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ , Robust standard errors in parentheses.

more likely to volunteer. We regress the volunteering decision on beliefs in a linear probability model to investigate this relation in the data. Figure 5 visualizes those effects in a local linear regression.

Table 4 shows the results, which are in sharp contrast to Hypothesis 4 and the theoretical predictions from the Volunteer’s Dilemma. There exists a strong positive correlation between beliefs and the volunteering decision. In other words, subjects volunteer *more* if they believe that it is more likely that

there is another volunteer in their team. An increase in the belief by one percentage point increases the probability of volunteering by 0.5 percentage points (Model 1 in Table 4).

The tendency to contribute to a public good conditional on the contributions of others has been documented in other public good games (see, for instance, Rapoport, 1985; Fischbacher et al., 2001). However, compared to a classical public good game, the payoff remains the same when the number of volunteers increases in our setup. Clearly, this connection between beliefs and actions is not individually rational from a monetary perspective. However, it is well in line with results from the literature on (social or self-) image concerns (for example, Bénabou & Tirole, 2006) or the vast literature on descriptive norms as well as on peer effects (see, for example, Cornelissen et al., 2017, for a recent study on peer effects at the workplace). We interpret this effect as a normative driver for the outcomes of our field experiments. Compared to a standard volunteers dilemma in the laboratory, it is reasonable to assume that subjects in the workplace environment want to comply with the decision of their coworkers, that is, they do not want to be seen (or see themselves) as unsocial or lazy by acting differently than what they believe is the ‘correct’ thing to do. While the situation is arguably still very anonymous towards other workers, social image concerns could still apply towards the workplace or the experimenter.

Note that this finding has to be considered jointly with the results that beliefs are, on average, lower in the small group compared to the two larger group sizes (Result 5). That is, while participants, on average, have lower beliefs about the volunteering of their group members, those with higher beliefs react more strongly to the beliefs compared to those with similar beliefs in the larger groups. Therefore, we find differences in the beliefs but no differences in volunteering rates across group sizes

**Result 6.** Workers are “conditional volunteers” and volunteer more if they believe others volunteer as well.

While Result 6 offers some insights into the driving forces of volunteering in our setup, it is essential to note that beliefs are not exogenous in this setting. We therefore consider further factors.

#### 4.6. Replication and robustness of our results

Our results on volunteering behavior and beliefs are surprisingly different from previous results from lab and field studies on the Volunteer’s Dilemma. Thus, our first step is to confirm the validity of our results in a replication study. We find qualitatively and quantitatively equivalent results in another experiment with  $N = 2,733$  participants on the same platform: Group size does not correlate with volunteering behavior. Volunteering rates in the replication range from 0.522 to 0.545, compared to the 0.514 to 0.552 reported in this study. Section C.1 in the Appendix provides a detailed discussion of the other experiment demonstrating the robustness of our findings.

In addition, we provide extensive randomization checks in Appendix C. We find that randomization worked very well, and the treatments are balanced according to several observable participant characteristics (see Table B.5 and Figure C.1 in the Appendix C).

Importantly, we find strong evidence that participants are well aware of their group size. First, while volunteer rates do not differ across treatments, beliefs are sensitive to group size, suggesting that participants were aware of it. Second, we explicitly asked participants at the end of the experiment if they remembered their group size, and 93 % did recall it correctly (see Table C.2 in Appendix C). This demonstrates that group size was salient and that participants were aware of it.

Overall, workers understood the incentives involved, and thus our results are not driven by a general misunderstanding of the task. In a further experiment discussed in Section 4.7 we included control questions to ensure that there are no misunderstandings of the incentives to volunteer. Also we informed participants about the correct answers. In general, understanding was high and around 80–85% of subjects answered the key control questions correctly. Crucially, also in the part

of the study resembling our main treatments, we find high volunteering rates and no effect on group size.

To ensure that our results regarding beliefs are robust, we thoroughly analyze additional measures we obtained in the questionnaire on economic preferences such as reciprocity, risk preferences, altruism, and so on. (see Table C.3 and Section D in the Appendix). We find several main effects, for example, risk preferences negatively correlate with volunteering, but no substantial interaction that would come close to the explanatory power of measured beliefs.

While there is a significant main effect of reciprocity on volunteering, this effect does not explain the main findings that volunteering is insensitive to the group size. Table D.8 in the Appendix shows that there is no significant interaction effect between treatment and reciprocity on volunteering. However, this makes intuitive sense: If someone is certain that others will volunteer, equalizing the payoffs by volunteering is costly and does not add to anyone's welfare. This speaks for the robustness of the explanation of beliefs as a driver of normative behavior.

Finally, we would like to reiterate that our baseline treatment allows us to rule out the possibility that participants did not understand the strategic incentives in the experiment or that volunteering was not costly enough. As summarized in Result 1, participants volunteer more when their payoff is independent of the decision of other team members ( $N = 1$ ) and less when they are not paid for it. Thus, participants understood the strategic incentives of the experiment and that volunteering is costly, further strengthening the validity of our main results.

#### 4.7. Purpose of the task

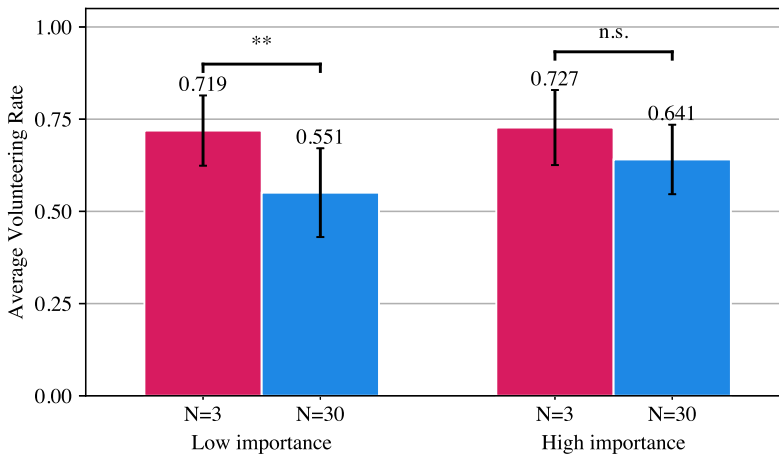
This section provides a potential explanation for the stark contrast between our robust results and previous (lab) experimental studies. One distinct feature in our work environment is that the task has a purpose beyond simply ensuring payments for the decision-makers and taking part in an organizational context. It might change the decision frame of decision-makers (Northcraft & Tenbrunsel, 2011) from an abstract strategic environment to an organizational context. As results by Kim & Murnighan (1997) suggest, measuring the intentions to volunteer within organizational scenarios may “dampen the effects of self-interest and other structural variables”. Thus, instead of perceiving the volunteering choice as a strategic situation in which the group size would be crucial, workers might perceive the game as non-strategic and focus more on the purpose of the task and maybe their individual costs.

Using an additional experiment with 338 subjects and focusing only on small and medium group sizes (3 and 30 workers), we actively vary the purpose of the volunteering choice. In the high-importance condition, the instructions are nearly identical to those in our main study. In the main study, people are informed at the beginning of the study that the task per se is meaningful as the ratings in the first part and that of one volunteer will be used in a scientific study but in the additional experiment we again emphasize this right before the volunteering decision.<sup>19</sup> In the low-importance condition, we informed the participants that their work in the volunteering task would not be used, thus “downgrading” the importance of the work to a mere financial bonus task and opening the decision further to strategic considerations.<sup>20</sup> We provide a screenshot of the decision screens and the translation in Appendix E.

In line with our previous results, we find no significant group size effect in the high-importance setting (see Figure 6, 72.7% vs. 64.1%,  $\chi^2$  test,  $p > 0.1$ ). However, we find a significant group size effect for the low-importance treatment where volunteering under the medium group size is significantly

<sup>19</sup>Note that the ratings of workers in the first part and those of one volunteer per group will be used in follow-up studies by Álvarez-Benjumea & Winter (2018, 2020) and Álvarez-Benjumea (2020).

<sup>20</sup>As a further difference to our main studies, we included control questions which might further emphasize that the data of the study is used for scientific purposes (high importance) or not (low importance).



**Figure 6.** The average volunteering rate across the different group sizes and with varying task importance. The error bars represent 95% confidence intervals

lower (71.9% vs. 55.1%,  $\chi^2$  test,  $p < 0.05$ ). Taken together, we find no level effect due to importance, that is, importance does not generally shift the willingness to volunteer.

Our results suggest that the importance or purpose of the task might shift the perception of the situation, with more strategic motives only entering the decision-making process under the low-importance tasks.

## 5. Discussion and concluding remarks

In this paper, we study volunteering at the workplace. While volunteering as an allocation mechanism in work environments is now widespread, and the industry strongly advocates its adoption, economic analysis, and empirical results suggest that two factors might impede volunteering. First, the incentives create a coordination problem, that is, who should volunteer, and second, workers prefer others to do the work, often leading to a situation similar to the Volunteer's Dilemma. In particular, our game-theoretical model and past research predict a higher diffusion of responsibility in larger groups, that is, lower volunteering rates.

Following the long standing tradition of using experimental methods to understand business and organizational problems (see, for instance, Rapoport & Zwick, 2002; Camerer & Weber, 2013), we report the results of a field experiment with more than 2,000 workers conducted in an online labor market. In our experiment, our workers individually worked on a standard classification task, and were then asked to volunteer for a similar task to secure a bonus for all their team workers. We exogenously vary the team size and thus study the causal effect of team sizes on volunteering behavior. Additionally, we elicit workers' beliefs about their co-workers' volunteering probability.

We find that workers react to the strategic situation of the Volunteer's Dilemma and shifts in the incentive structure. Workers volunteer much more if they are alone instead of embedded in a group and thus a Volunteer's Dilemma. They also volunteer much less if they are not paid for the bonus task. Furthermore, in line with our theoretical predictions, workers with lower subjective costs are more likely to volunteer.

In stark contrast to previous results that study volunteering outside of the remote work context, we find no effect of the group size on volunteering behavior. On average, about 53% of workers choose to volunteer across all our main treatments.

We identify workers' beliefs about the volunteering decision of their team members as a key driver of volunteering in our settings. Workers who believe that it is more likely that at least one other worker

volunteers are *more* likely to volunteer themselves. This *conditional volunteering* behavior is puzzling from a purely material perspective. However, it aligns with findings by Rapoport et al. (1989) and with results from descriptive norms (Bicchieri, 2005). If workers perceive volunteering as a stronger descriptive norm, indicated by higher beliefs, they might be more inclined to follow these norms. A similar explanation can stem from image concerns. Workers do not want to be perceived as selfish by others or themselves for not volunteering if they believe that most other workers are volunteering. While this offers some insights into the driving forces of volunteering in our setup, it is essential to note that beliefs are not exogenous in this setting and we cannot rule out that other confounding factors simultaneously influence beliefs and volunteering choices. Nonetheless, it would be an interesting path for future research to study how beliefs influence the volunteering choice explicitly in a treatment design.

In contrast to previous settings in the Volunteer's Dilemma where a group-size effect is usually observed, our workers make decisions in an organizational context with a task that has a purpose beyond simply providing financial incentives. This purpose and the organizational setting might change the perception of the game for workers, putting them in a "work frame" (Northcraft & Tenbrunsel, 2011) and thus dampening strategic considerations like group size. Similar arguments were used in an earlier study by Kim & Murnighan (1997) in which also no group size effects on intended volunteering are reported in a (framed) workplace scenario. Using additional experiments, we explicitly shut down the purpose of the task by telling workers that the work they do in the additional task will not be used for anything else than determining the bonus. Interestingly, in this setting, we observe a group size effect with workers in larger groups volunteering significantly less. Thus, having a purpose for the task seems to crowd out some of the strategic considerations often assumed when modeling the volunteer decision theoretically (see also Section 3), or considering it in more abstract lab experiments. These differences should be analysed in more detail in future studies and our results can therefore provide a starting point for an encompassing model of volunteering at the workplace.

Our results suggest that volunteering at the workplace is distinctively different from other volunteering situations. Thereby, we contribute to a growing body of literature studying the scope conditions of group-size effects: Weimann et al. (2019) show that groups of up to 100 members successfully provide public goods, even though free-riding incentives are exceptionally high. The same observation has recently been found in peer punishment (Carpenter, 2007) and third-party punishment experiments (Kamei, 2020).

Those findings have several implications for firms' organizational structure or, more generally, social groups that rely on volunteering. We deliberately set up the work environment so that workers are not informed whether others worked before them, which allows us to measure their individual willingness to volunteer. However, depending on the exact organizational structure, our results might be good or bad news for volunteering at the workplace. Interpreted strictly within the context of the Volunteer's Dilemma, the high level of volunteering is wasteful and inefficient. Especially in work settings where it is not possible for a manager to prevent this over-provision, the allocation mechanism has adverse welfare effects. Indeed, these inefficiencies are discussed in open source software development (McConnell, 1999; Kenwood, 2001). It is especially relevant in workplaces with limited communication between workers or with little oversight by a manager, like in remote work contexts that have become more popular in recent years. Here, managers should focus on smaller team sizes to prevent too much volunteering and reduce inefficiencies.

For organizations that cannot prevent over-provision and only care about finding at least one volunteer, that is, who only care about producing the good, our results are good news. Contrary to what theory or intuition would suggest, giving workers the freedom to self-organize does not hurt commitment. Using larger team sizes might be desirable from the perspective of a manager. It is true even

without oversight and without any disciplinary measures. Our findings might thus relate to high volunteering rates in other contexts, such as writing open-source code or Wikipedia articles, even though the number of potential volunteers is sometimes very large.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/eec.2024.13>.

**Acknowledgements.** We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 395336584 and the ZEW - Leibniz Zentrum für Europäische Wirtschaftsforschung. Ethical approval was obtained from the German Association for Experimental Economic Research e.V. and can be accessed under <https://gfew.de/ethik/Ft9eR5SK>. The replication material for the study is available at <https://doi.org/10.17605/OSF.IO/PRWCS>.

**Statements and declarations.** The authors declare that they have no competing financial or non-financial interests.

## References

- Álvarez-Benjumea, A. (2020). Exposition to xenophobic content and support for right-wing populism: the asymmetric role of gender. *Social Science Research*, 92, 102480.
- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34 (3), 223–237.
- Álvarez-Benjumea, A., & Winter, F. (2020). The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks. *Proceedings of the National Academy of Sciences*, PNAS, Vol. 117, pp. 22800–22804.
- Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107, 714–747.
- Barron, G., & Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18, 507–520.
- Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96, 1652–1678.
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- Bick, A., Blandin, A., & Mertens, K. (2021). Work from home before and after the Covid-19 outbreak. Available at SSRN 3786142.
- Boosey, L., Brookins, P., & Ryvkin, D. (2017). Contests with group size uncertainty: Experimental evidence. *Games and Economic Behavior*, 105, 212–229.
- Borchert, K., Hirth, M., Kummer, M., Laitenberger, U., Slivko, O., & Viete, S. (2023). Unemployment and online labor. *Management Information Systems Quarterly*, 47, 771–802.
- Camerer, C., & Weber, R. (2013). Experimental organizational economics. *The Handbook of Organizational economics* (pp. 213–262). Princeton: Princeton University Press.
- Campos-Mercade, P. (2021). The Volunteer's Dilemma explains the bystander effect. *Journal of Economic Behavior & Organization*, 186, 646–661.
- Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60, 31–51.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. *American Economic Review*, 107, 425–456.
- Danz, D., Vesterlund, L., & Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112, 2851–2883.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383.
- Diekmann, A. (1985). Volunteer's Dilemma. *Journal of Conflict Resolution*, 29, 605–610.
- Diekmann, A. (1986). Volunteer's Dilemma. A social trap without a dominant strategy and some empirical results. In *Paradoxical Effects of Social Behavior*, (pp. 187–197) Physica-Verlag HD.
- Diekmann, A. (1993). Cooperation in an asymmetric Volunteer's Dilemma game theory and experimental evidence. *International Journal of Game Theory*, 22, 75–85.
- Difallah, D. E., Catasta M., Demartini G., Ipeirotis P. G., & Cudré-Mauroux P. (2015). The dynamics of micro-task crowdsourcing: the case of Amazon MTurk, Proceedings of the 24th International Conference on World Wide Web, Geneva: International World Wide Web Conferences Steering Committee, in pp. 238–247.
- Enke, B., & Graeber, T. (2023). Cognitive uncertainty. *Quarterly Journal of Economics*, 138 (4), 2021–2064.
- Enke, B., Graeber, T., Oprea, R. (2023). Complexity and time. Tech. rep., National Bureau of Economic Research.
- Erev, I., & Rapoport, A. (1990). Provision of step-level public goods: the sequential contribution mechanism. *Journal of Conflict Resolution*, 34, 401–425.



- Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2023). The preference survey module: a validated instrument for measuring risk, time, and social preferences. *Management Science*, 69, 1935–1950.
- Farrell, D., Greig, F. (2017). The online platform economy: has growth peaked? Tech. rep., JP Morgan Chase Institute.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71, 397–404.
- Franzen, A. (1995). Group size and one-shot collective action. *Rationality and Society*, 7, 183–200.
- Goeree, J. K., Holt, C. A., & Smith, A. M. (2017). An experimental examination of the Volunteer's Dilemma. *Games and Economic Behavior*, 102, 303–315.
- Hans Böckler Stiftung (2021). Studien zu Home Office und Mobiler Arbeit. Tech. rep., Hans Boekler Stiftung.
- Hoda, R., Salleh, N., & Grundy, J. (2018). The rise and evolution of agile software development. *IEEE Software*, 35, 58–63.
- Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80, 984–1001.
- Jain, A., Sarma A. D., Parameswaran A., & Widom J.. (2017). Understanding workers, developing effective tasks, and enhancing marketplace dynamics: a study of a large crowdsourcing marketplace. *Proceedings of the VLDB Endowment*, Vol. 10. pp. 829–840.
- Johnson, J. P. (2004). Open source software: private provision of a public good. *Journal of Economics & Management Strategy*, 11, 637–662.
- Kamei, K. (2020). Group size effect and over-punishment in the case of third party enforcement of social norms. *Journal of Economic Behavior & Organization*, 175, 395–412.
- Kenwood, C. A. (2001). A business case study of open source software. Tech. rep., MITRE CORP BEDFORD MA.
- Kim, J. W., & Murnighan, J. K. (1997). The effects of connectedness and self interest in the organizational volunteer dilemma. *International Journal of Conflict Management*, 8, 32–51.
- Kopányi-Peuker, A. (2019). Yes, I'll do it: a large-scale experiment on the Volunteer's Dilemma. *Journal of Behavioral and Experimental Economics*, 80, 211–218.
- Kugler, T., Rapoport, A., & Pazy, A. (2010). Public good provision in inter-team conflicts: effects of asymmetry and profit-sharing rule. *Journal of Behavioral Decision Making*, 23, 421–438.
- Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, 89, 308–324.
- Lee, S., Levin, D., Gopalakrishnan, V., & Bhattacharjee, B. (2007). Backbone construction in selfish wireless networks. *AMC SIGMETRICS Performance Evaluation Review*, 35, 121–132.
- McConnell, S. (1999). Open-source methodology: ready for prime time? *IEEE Software*, 16, 6–8.
- Murnighan, J. K., Kim, J. W., & Metzger, A. R. (1993). The volunteer dilemma. *Administrative Science Quarterly*, 38, 515–538.
- Northcraft, G. B., & Tenbrunsel, A. E. (2011). Effective matrices, decision frames, and cooperation in volunteer dilemmas: a theoretical perspective on academic peer review. *Organization Science*, 22, 1277–1285.
- Otsubo, H., & Rapoport, A. (2008). Dynamic Volunteer's Dilemma over a finite horizon: an experimental study. *Journal of Conflict Resolution*, 52, 961–984.
- Przepiorka, W., & Berger, J. (2016). The sanctioning dilemma: a quasi-experiment on social norm enforcement in the train. *European Sociological Review*, 32, 439–451.
- Rapoport, A. (1985). Provision of public goods and the MCS experimental paradigm. *American Political Science Review*, 79, 148–155.
- Rapoport, A., & Amaldoss W. (2013). Social dilemmas embedded in between-group competitions: effects of contest and distribution rules. In Foddy Margaret, Smithson Michael, Schneider Sherry, Hogg & Michael A. (Eds.) *Resolving Social Dilemmas* (pp. 67–85). Psychology Press.
- Rapoport, A., Bornstein, G., & Erev, I. (1989). Intergroup competition for public goods: effects of unequal resources and relative group size. *Journal of Personality and Social Psychology*, 56, 748.
- Rapoport, A., & Suleiman, R. (1993). Incremental contribution in step-level public goods games with asymmetric players. *Organizational Behavior and Human Decision processes*, 55, 171–194.
- Rapoport, A., and Zwick, R. (2002). *Experimental Business Research*, Springer.
- Weesie, J. (1994). Incomplete information and timing in the Volunteer's Dilemma: a comparison of four models. *Journal of Conflict Resolution*, 38, 557–585.
- Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H., & Keser, C. (2019). Public good provision by large groups—the logic of collective action revisited. *European Economic Review*, 118, 348–363.
- Zhang, X. M., & Zhu, F. (2011). Group size and incentives to contribute: a natural experiment at Chinese Wikipedia. *American Economic Review*, 101, 1601–1615.