Journal of Business Economics (2025) 95:237–256
https://doi.org/10.1007/s11573-024-01214-7

ORIGINAL PAPER

Re-examining the effect of wage delegation: a replication study of Charness et al. (2012)

Hendrik Niehoff[1] · **Philipp Schreck**[1]

Accepted: 31 October 2024 / Published online: 4 December 2024

Abstract
Experimental research has explored the effects of wage delegation on employee performance, with the pioneering attempt by (Charness et al., Am Econ Rev 102:2358–2379, 2012) in "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game", published in the American Economic Review. We conducted a replication of their experiment for two main reasons: first, to validate the original findings that have influenced subsequent research, and second, to examine whether these results hold in contexts with reduced social proximity between workers and employers, a relevant issue due to the rise of remote work. Our online experiment, involving 410 participants, followed the original study's design but used a different sample. We successfully replicated the main finding that wage delegation increases employee effort, though the effect was smaller and largely due to employees granting themselves higher wages. These results support the notion that with decreasing social proximity, formal controls increase in importance.

Keywords Wage delegation · Replication study · Performance incentives · Employee empowerment · Intrinsic motivation

JEL Classification C91 · J31 · J33 · J41 · M52

1 Introduction

A growing body of evidence originating from the field of employee empowerment suggests that increasing the degree of autonomy for workers can foster a greater sense of responsibility and job satisfaction (Maynard et al. 2012; Yin et al. 2019).

✉ Hendrik Niehoff
hendrik.niehoff@wiwi.uni-halle.de

Philipp Schreck
philipp.schreck@wiwi.uni-halle.de

[1] School of Law and Economics, Martin-Luther-University Halle-Wittenberg, Große Steinstr. 73, 06108 Halle (Saale), Germany

By allowing employees more freedom, their motivation can naturally increase, leading to improved productivity. In particular, it has been argued that the democratization of organizations may increase employee satisfaction, effort and performance on multiple dimensions (Harrison and Freeman 2004). Importantly, employee empowerment may imply the delegation of wage decisions to employees themselves with the goal of increasing employee autonomy and satisfaction (Mellizo et al. 2014). In line with this notion a wide range of businesses are practicing this autonomy-encouraging approach already, including companies such as Semco from Brazil, Skyline from the United States, or the Spanish company Claravision (Semler 2007; Tuna 2008).

Over the years, various experimental studies have investigated the effects of wage delegation on employee performance, with the pioneering attempt to study this relationship by Charness et al. (2012) in "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game", published in the American Economic Review. This landmark study aimed to analyze the outcomes of wage delegation in a principal-agent setting via an incentivized lab experiment. The results of this experiment provide strong evidence of an increase in both wages and worker effort as a result of delegating wage determination. The authors attribute this effect primarily to employees' increased sense of responsibility, rather than to positive reciprocity toward their employers (Charness et al. 2012).

In the current paper, we report the results of a study replicating the experiment of Charness et al. (2012). There are two main reasons for our replication study. The first is that the insights from the study by Charness et al. (2012) have served as a foundation for numerous follow-up studies, each varying in design and empirical approach. For instance, Jeworrek and Mertins (2019) conducted a natural field experiment that validated the positive effect of wage delegation on employee performance in the field. Also other studies, such as those by Mellizo et al. (2014), Charness et al. (2016), Franke et al. (2016), and Faillo and Piovanelli (2017), have built upon the insights of Charness et al. (2012) utilizing different experimental designs to further explore the effects of wage delegation. More recently, Brück et al. (2021) shifted the focus away from the relationship between wage delegation and employee performance and explored the impact on employee creativity. This array of diverse articles underscores the significance and potential importance of replicating the study by Charness et al. (2012).

The second main reason for our effort to replicate Charness et al. (2012) is that we want to explore some of the study's boundary conditions. The original study was conducted in a behavioral laboratory, which is an apt setting for mimicking situations of social proximity as they prevail in organizations. In such settings, interaction partners can develop a shared understanding of a situation, and social norms such as trust and reciprocity may evolve. The original study argued that participants in the role of employees may have increased effort in response to delegation due to an enhanced feeling of responsibility towards the employee (Charness et al. 2012: 2360).

However, experimental research has shown that an increased social distance between interaction partners strongly affects how they think and feel about the interaction (Charness et al. 2007; Schmelz and Ziegelmeyer 2020), and that this

distance might have detrimental effects on workers performance (e.g. Cramton 2001; Cramton and Webber 2005; Burbano and Chiles 2022). These findings raise the question whether the original results also hold in settings where people interact across larger distances, as is the case when they work from home. Given the current trend of growing physical distances between companies and employees working remotely (Bloom et al. 2023), this is a highly relevant question. We thus use the design of the original experiment and seek to replicate it using an online participant pool. Conducting experiments online vs. in the lab captures "meaningful differences between working from home and working at the office arrangements as online subjects enjoy greater anonymity than lab subjects, they interact in a less constrained environment than the laboratory, and there is a larger physically oriented social distance between them." (Schmelz and Ziegelmeyer 2020: 958).

Replication studies play an indispensable role in economics and management accounting, though they represent only a modest proportion of all recent academic publications, with a large number of highly cited papers yet to be replicated (Albers 2014; Hensel 2019; Mueller-Langer et al. 2019; Brüggen et al. 2021). This limited prevalence can be attributed to hurdles such as issues around data and code access and a lack of strong incentives to publish these types of studies (Ditzen and Elhorst 2022). Furthermore, the literature lacks a universally agreed-upon demarcation for the concepts of reproducibility, replicability, and robustness and moreover different varieties of replication forms exist (Duvendack et al. 2017; Köhler and Cortina 2021).

Nevertheless, the relevance of replication studies in experimental economics has gained increasing prominence, particularly due to the so-called replication crisis in psychology, which has sparked these discussions (Maniadis et al. 2015; Open Science Collaboration 2015). Past replications have often yielded results that were inconsistent with the original studies (Chang and Li 2022). For instance, the findings of a large-scale replication project conducted by Camerer et al. (2016) demonstrated that only 11 out of 18 replicated studies (61%) published in the American Economic Review and the Quarterly Journal of Economics, covering the period from 2011 to 2014, showed a significant effect in the same direction as the original studies.

In light of the importance of the Charness et al. (2012) study and the widely-accepted need for replications of influential empirical results, our study can be best described as a "scientific replication," in terms of the classification of Hamermesh (2007): It draws a sample from a different population, but uses the same estimation strategy as the original study. Because we conduct our study online, rather than in an on-site behavioral laboratory, both the population and the sample are different from the original study. However, the specification remained the same, as the original experiment was precisely reprogrammed and not modified.

Based on their study results, Charness et al. (2012) concluded that wage delegation significantly enhances workers performance and increased earnings for both firms and workers. Interestingly, Charness et al. (2012) found that the increased effort was not primarily motivated by positive reciprocity; rather, Charness et al. (2012) postulated that it might have reflected a heightened sense of responsibility that workers felt towards their outcomes.

The results of our replication study confirm that wage delegation does indeed increase worker performance. However, there are important differences in the results of the original and the replication study. Wage delegation only appears to positively affect worker effort in one-shot interactions (stranger matching protocols). In repeated interactions (partner matching protocols), the increased effort appears to be driven solely by higher wages and the higher efforts desired by employers. Additionally, the results indicate that in the stranger scenario, the deliberate decision by the employer not to engage in wage delegation negatively impacts worker productivity levels. A similar negative reaction to non-delegation was observed by Franke et al. (2016), who investigated varying degrees of workers' involvement in the wage-setting process using a laboratory gift-exchange game. Finally, in contrast to the original study, our data do not rule out the possibility that higher worker effort levels in response to wage delegation are partly due to positive reciprocity. This result also contrasts with subsequent studies in the field of wage delegation, which, like Charness et al. (2012), were able to rule out the effect of positive reciprocity as an explanation for the higher effort under delegation (Jeworrek and Mertins 2019).

2 Study design

2.1 Research question and design of the original study

The fundamental question that Charness et al. (2012) asked is this: How do employers and employees alter their behavior when presented with the option to delegate wage decisions? To answer this question, Charness et al. (2012) conducted a laboratory experiment in Granada, Spain, involving 236 student participants. Participants were assigned fixed roles as either a firm or a worker, and performed a stated effort task in groups of two. Workers could provide costly effort which would increase firm profit; the firm, in turn, paid out a fixed wage to the worker. More formally, firm and worker payoffs were determined by the following profit functions:

$$\Pi_F = (240 - w) \times e \tag{1}$$

$$\Pi_W = w - c(e) - 20 \tag{2}$$

where F denotes the firm, W the worker, e is the worker's effort level, and w the wage with $w \in [20, 21, \ldots 119, 120]$. $c(e)$ represents the cost of effort (increasing in e). Table 1 lists the feasible effort levels and corresponding costs.

In essence, the setting represents a gift-exchange game: The worker needs to provide (costly) effort to increase the firm's payoff, while the firm needs to provide a (costly) wage to increase the worker's payoff. To introduce the idea of wage delegation, employers in some treatments could choose whether to set the employee's wage by themselves or whether to allow the employee to determine his own wage.

The experiment consisted of 15 periods and 2 initial test rounds. At the start of each period, participants in the role of the firm (the employer) had to specify an expected, non-binding effort from the worker, and decided whether to set the

Table 1 Effort levels and costs of effort

Effort e	Cost $c(e)$
0.1	0
0.2	1
0.3	2
0.4	4
0.5	6
0.6	8
0.7	10
0.8	12
0.9	15
1.0	18

worker's wage or whether to delegate this decision to the worker. The worker was first informed of the expected effort level and the employer's wage-setting decision. If the employer chose to delegate, the worker could set his own wage. If, in contrast, the employer chose to set the wage by herself, the worker was informed of the wage. Finally, the worker had to state the level at which he was willing to exert effort.

The experiment was implemented in a $2 \times 2 + 1$ factorial design, resulting in five treatments. Participants were unique to each treatment and session, with two sessions per treatment and between 22 and 24 individuals per session. The wage setting mechanism (delegation yes/no) and the matching procedure (partner v. stranger) were manipulated at two levels, each. In the Partner treatments, firms and workers were anonymously paired for all periods, while in the Stranger treatments, pairs were randomly rematched each period. In the Control treatments, the firm had to set wages without having the option to delegate this decision to the worker. The fifth treatment, the Dictator treatment, used a stranger matching protocol and implemented delegation as the norm: here, only the worker was able to set the wage.

The experimental design allows to measure the effect of delegation on wages, effort levels, and profits, depending on whether delegation was (endogenously) chosen by the firm, or exogenously set in the specific treatment.

2.2 Design of the replication study

The goal of our study is to determine whether the results of Charness et al. (2012) are replicable which would lend further support to its validity and generalizability. Accordingly, we chose an experimental design that would come as close as possible to the original study. In terms of the classification of Hamermesh (2007), our study can be best described as a "scientific replication:" It draws a sample from a different population, but uses the same estimation strategy as the original study. In particular, the specification remained unchanged, as the original experiment was precisely reprogrammed and not modified. As summarized in Table 2, our study differs from the original study in four aspects that are not related to the specification.

Table 2 Differences in the designs of the original and the replication study

Categories	Charness et al. (2012)	Replication study
Sample size	236	410
Experimental method	Laboratory experiment	Online experiment
Sample	Students from the University of Granada	Prolific Pool (UK)
Number of periods (without training periods)	15 periods	10 periods

In the original study by Charness et al. (2012), a total of 236 participants were involved; however, the authors provided no rationale for this specific sample size. To determine the necessary sample size for our replication study, we performed an ex-ante power analysis. The first step in this analysis involved determining the effect size for each effect mentioned by Charness et al. (2012). This was achieved by initially calculating the pooled standard deviation for each effect using the original data. This pooled standard deviation, in combination with the group means reported in the paper, enabled the calculation of Cohen's d for each effect. In the next step Cohen's d was used to determine the necessary sample size for each effect in G*Power (Faul et al. 2007) with the alpha error set at the conventional threshold of 0.05 and the power set to 95% in order to detect the original effect size at the 5% significance level. The analysis yielded a required sample size of 80 for each treatment. In our effort to reach this minimum size, we ended up with a total of 410 participants.

We recruited the participants of our experiment using Prolific, a popular subjects pool for online academic studies, which offers a large and diverse base of subjects (Peer et al. 2017; Palan and Schitter 2018). The original experiment was a computerized laboratory experiment, programmed in z-Tree (Fischbacher 2007). We used the original z-Tree files made available online by Charness et al. (2012) to accurately replicate the experiment using the Lioness platform (Giamattei et al. 2020). This replication ensured that all aspects of the experiment, including the screens and other elements, were identical to the original computerized experiment. The experiment was conducted on four consecutive days with at least two sessions per treatment. In line with the design from Charness et al. (2012), no participant was able to participate in more than one session or treatment. In response to dropout rates of more than 70% in our pilot study, we decided to slightly adjust the number of rounds in our replication study. Attrition in online experiments is a well-known issue (Arechar et al. 2018; Giamattei et al. 2020), but given the specific study design, we had to deal with an unusually high dropout rate. In treatments with a partner matching process, each participant who dropped out in any of the 15 periods, left an unmatched partner who had to leave the experiment prematurely. In treatments with a stranger matching process, participants were matched within cohorts. These had to be large enough to allow for stranger matching, and small enough to avoid excessively long waiting times, which would have resulted in higher dropout rates. Hence, a single participant's dropout would end the experiment for the entire cohort because others who were intended to be matched with this specific participant in later stages of an

experimental session, could not be matched anymore. To mitigate these problems of attrition, we decided to limit our experiment to 12 rounds, including 2 training rounds, and chose a cohort size of four. We thus implemented an imperfect stranger matching where no participant would be matched with the same partner in two consecutive rounds.[1] Participants were informed about being rematched after each round, but were unaware of the size of the cohort. We can hence assume that participants could not identify repeated interaction and reputation building was not possible. Taken together, these measures led to a decrease in dropout rates to 45%.

3 Results of the original study and the replication study

3.1 Delegation effects on effort and earnings

Table 3 provides mean values of the main variables across the five treatments for the original study and the replication study. For each of the two treatments where delegation was possible, the mean values are presented in two separate columns, depending on whether the wage decision was made by the employer or delegated to the employee in a given period (Partner nondelegation v. Partner delegation; Stranger nondelegation v. Stranger delegation).

First we turn to the results of Charness et al. (2012). In the Partner treatment, wages and actual effort levels were higher when firms delegated (Partner delegation) than when they did not delegate (Partner nondelegation).[2] The comparison of the Partner nondelegation case with the Control Partners treatment, where no option to delegate existed, showed little difference in wage and effort levels. This lack of difference suggests that nondelegation—where it was possible—had no demotivating effect *per se*.

Under the stranger matching protocol, results were very similar. Delegation resulted in workers selecting higher wages and more effort. That is, positive effects of delegation persist, even if firms and workers cannot act strategically. However, in the Stranger treatment, workers exerted less effort compared to the Partner treatment, highlighting the role of enduring employment relationships in influencing worker behavior. These results suggested that workers may increase their effort when they consider long-term strategic factors.

Charness et al. (2012) suggested that the increased effort in delegated scenarios might have been due to two distinct factors: First, it could have stem from workers' positive response to delegation or from their negative reaction to intentional nondelegation. Second, it could have been due to higher wages that workers set for themselves if they have the chance to do so. To disentangle these effects, for each of the treatments where delegation was possible, Charness et al. (2012) regressed effort on

[1] The article by Charness et al. (2012) does not provide a definitive clarification on whether the stranger matching process used in the experiment is a perfect or imperfect stranger mechanism.

[2] If not stated differently, all differences discussed in this section are significant in a statistical sense ($p \leq 0.001$). To test for statistical differences, Charness et al. (2012) used two-tailed Wilcoxon signed-ranks tests.

Table 3 Summary of the five experimental treatments (Charness et al. 2012 and replication results)[a]

	Partners PND	Partners PD	Partners control	Strangers SND	Strangers SD	Stranger control	Dictator
Wage							
Original	76.79	114.70	75.95	45.60	117.85	51.24	113.05
Replication	69.68	106.06	70.86	58.21	104.34	58.29	92.21
Desired effort							
Original	0.78	0.81	0.72	0.61	0.67	0.59	0.68
Replication	0.69	0.68	0.68	0.71	0.74	0.69	0.69
Actual effort							
Original	0.43	0.78	0.46	0.15	0.34	0.17	0.36
Replication	0.47	0.65	0.51	0.42	0.63	0.47	0.51
Firm earnings							
Original	64.29	84.62	68.36	27.53	41.91	30.24	45.93
Replication	72.50	85.77	80.82	74.53	85.88	86.28	78.32
Worker earnings							
Original	51.09	84.15	50.01	24.93	93.81	30.35	88.96
Replication	43.46	76.30	43.92	32.62	75.36	35.87	73.40
Total earnings							
Original	115.38	168.77	118.37	52.46	135.72	60.59	134.88
Replication	115.96	162.07	124.74	107.15	161.24	122.15	151.72

[a]The abbreviations in each column include "P" or "S", denoting the partner and stranger treatments, and "D" or "ND," indicating whether the scenarios with delegation or nondelegation were examined. "Wage" refers to the average wage set by either the firm or the worker across all periods. "Desired effort" denotes the average effort requested by firms, while "actual effort" represents the effort workers are willing to provide. "Firm earnings" and "Worker earnings" indicate the average profit earned by both workers and firms based on a predefined profit function across all periods. "Total earnings" is the cumulative profit achieved on average by both firm and worker over all periods

a set of explanatory variables, including wage, desired effort, a delegation dummy (taking value 1 if the firm chose to delegate in a given period, and 0 otherwise), and period.[3] As the results reported in Table 4 reveal, even when controlling for wage effects, delegation significantly and positively influenced effort in both matching scenarios.

Finally, delegation benefited both firms and workers by leading to higher total earnings. Specifically, in the Partners delegation scenario, total earnings, and hence social efficiency, were 50% higher compared to the Partners nondelegation setting.

[3] As we will explain in more detail further below, Charness et al. (2012) estimated Generalized Least Squares (GLS) random effects models.

Table 4 GLS Random effects regression analysis, partners and strangers treatment (Charness et al. 2012 and Replication Study)[b]

	Partners		Strangers	
	(1)	(2)	(3)	(4)
Wage				
Original		0.006*** (0.000)		0.0011*** (0.0004)
Replication		0.003*** (0.0005)		0.002*** (0.0006)
Desired effort				
Original		0.095 (0.059)		0.040 (0.025)
Replication		0.378*** (0.0831)		0.186** (0.0808)
Delegation				
Original	0.282*** (0.036)	0.073** (0.033)	0.173*** (0.042)	0.100*** (0.043)
Replication	0.0968*** (0.0352)	0.0119 (0.0262)	0.181*** (0.0456)	0.0969** (0.042)
Heterogeneity				
Original	No	Yes	No	Yes
Replication	No	No	Yes	Yes
Constant				
Original	0.451*** (0.037)	– 0.049 (0.155)	0.150*** (0.010)	– 0.245*** (0.093)
Replication	0.716** (0.3096)	– 0.4605* (0.2442)	0.503 (0.3752)	0.262 (0.3408)
R^2				
Original	0.236	0.583	0.149	0.284
Replication	0.224	0.528	0.154	0.232
N				
Original	360	360	360	360
Replication	420	420	400	400

[b]Notes: Robust standard errors are in parentheses. Delegation takes the value 1 if and only if the firm delegated the wage decision. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Also the Big Five Inventory consisting of the traits, extraversion, agreeableness, conscientiousness, neuroticism and openness was used to control for heterogeneity with each characteristic being computed as an average of two questions (Rammstedt and John 2007)

Additionally, average worker earnings notably increased with delegation in the Strangers matching protocol. Based on these results, Charness et al. (2012: 2367) summarized their first main finding as follows:

RESULT 1: Delegating the wage decision enhances worker performance and increases the earnings of both firms and workers relative with the case where firms do not delegate.

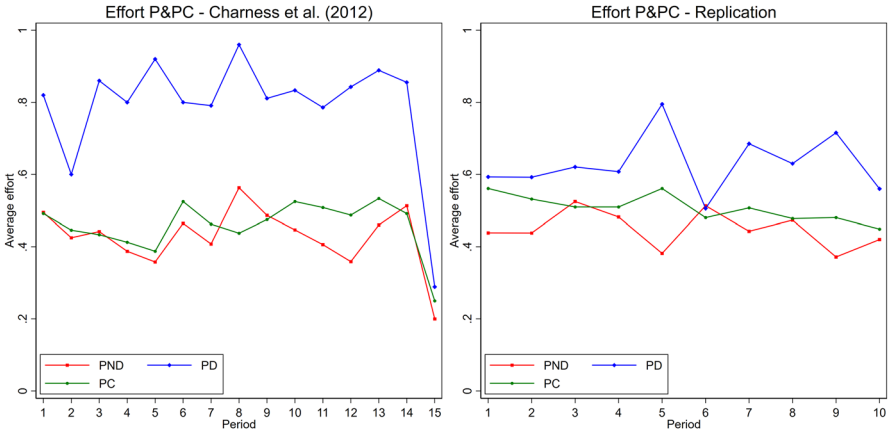

Fig. 1 Average effort over time, partner vs. control (Charness et al. 2012 and replication results)

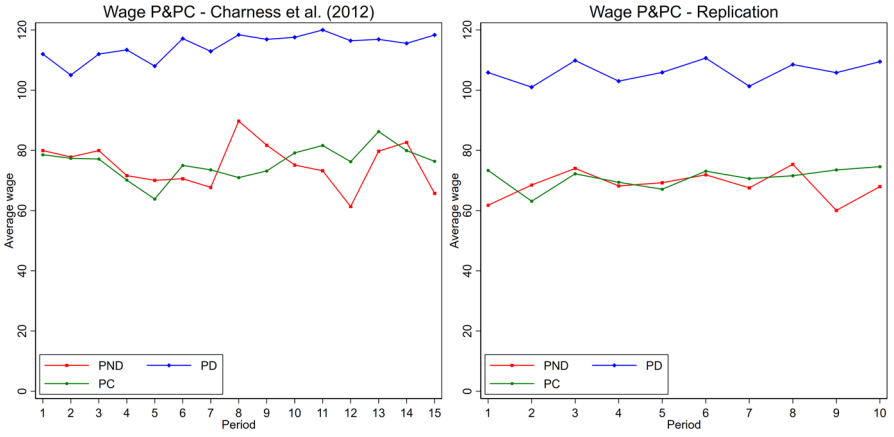

Fig. 2 Average wages over time, partner vs. control (Charness et al. 2012 and replication results)

We now turn to the analysis of the replication study results with a comparison of wages and effort *within* treatments, depending on whether the wage choice was delegated. In the Partners treatment, the results of the replication show that, similar to the original study, both wages and actual effort are higher when the firm delegates compared to the nondelegation case. A Wilcoxon matched-pairs signed-rank test confirms that these differences are significant ($Z = 4.036$, $p = 0.000$; $Z = 2.042$, $p = 0.041$, two-tailed tests for wages and effort respectively).[4] As Table 3, Figs. 1, and 2 reveal, the differences in the levels of effort are not as

[4] Interestingly, the original article (p. 2363) reports three statistics to test for the significance of two differences.

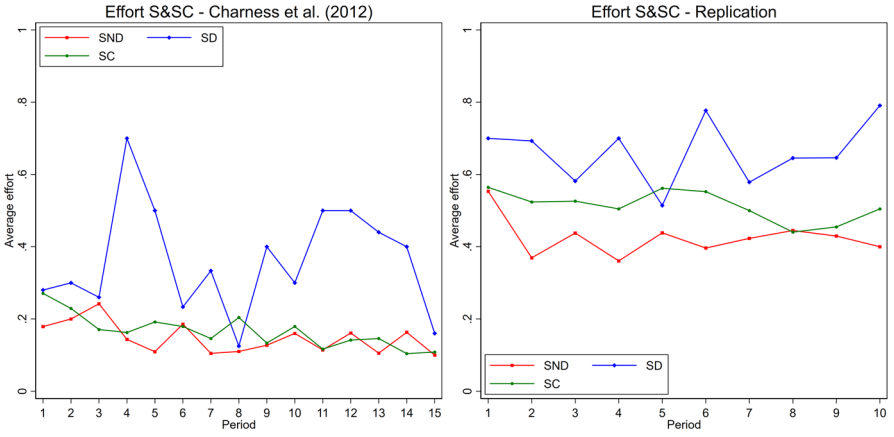

Fig. 3 Average effort over time, stranger vs. control (Charness et al. 2012 and replication results)

Fig. 4 Average wages over time, stranger vs. control (Charness et al. 2012 and replication results)

pronounced as in the original study, while differences in wage levels are very similar in both studies.

Additionally, the replication results for the partner treatment show only minimal (non-significant) differences in wage levels between the nondelegation scenario and the Control Partners treatment where delegation was not an option, as shown by a two-tailed Mann–Whitney test ($Z = -0.475$, $p = 0.635$). Effort levels also show a similar yet slightly significant trend ($Z = -1.639$, $p = 0.090$). This consistency in wage and effort responsiveness across both conditions reconfirms the idea that not delegating the wage decision does not necessarily lead to adverse outcomes.

In the Stranger treatments, the wage and actual effort values are higher in the delegation scenario compared to the nondelegation scenario. A Wilcoxon test confirms the significance of these differences, which is in line with the original results ($Z = 6.247$, $p = 0.000$; $Z = 3.712$, $p = 0.000$, for wages and effort, respectively). However, Table 3 and Fig. 3 show for the stranger treatment that the absolute effort values in the delegation, nondelegation and control cases differ from those of the original study. For example, in the nondelegation case, the average effort value is almost three times as high as in the original experiment (0.42 and 0.15).

As to the reasons for the increased efforts under delegation, we follow the procedure of Charness et al. (2012) and run a GLS regression. Here, the replication study yields remarkably different results in the Partner treatment than Charness et al. (2012). As Table 4 suggests, the differences in effort are mainly due to higher wages and desired effort, and not due to delegation per se in this treatment. In contrast, for the Stranger treatment, the replication study confirms Charness et al.'s (2012) findings, showing that delegation positively influences the level of effort chosen by workers in one-shot interactions.

Next, we consider earnings. Table 3 shows that in the replication study's Partner treatment under delegation, total earnings are approximately 40 percent higher than in the nondelegation case. This is broadly in line with the 50 percent increment reported by Charness et al. (2012). However, the distribution of these differences varies across the two studies. While in the original study delegation resulted in higher earnings for both, firms and employees, in the replication study only employees benefit from delegation. The differences in workers' earnings are statistically significant ($Z = 4.036$, $p = 0.000$, one-tailed Wilcoxon matched-pairs signed-rank tests). The difference in firm earnings, on the other hand, are not ($Z = 0.456$, $p = 0.648$, one-tailed Wilcoxon matched-pairs signed-rank test). This difference between the original and the replication studies is due to the fact that in our replication study, workers select higher wages if they can, but do not seem to exhibit the same sense of responsibility as in the original study in terms of exerting higher levels of effort.

In the Stranger treatment, we obtain very similar results as Charness et al. (2012). As Table 3 shows, the average firm earnings are higher in the case of delegation compared to nondelegation (85.88 versus 74.53). Even though this difference is larger than in the original study, it is not significant, mirroring the findings reported by Charness et al. (2012) ($Z = 0.608$, $p = 0.562$, two-tailed test).

3.2 Explaining the delegation effects: positive vs. negative reactions and positive reciprocity vs. sense of responsibility

To investigate any negative response to intentional nondelegation, Charness et al. (2012) compared those cases in which firms chose not to delegate (PND and SND) with the control treatments, in which firms did not have the choice to delegate (PC v. SC, respectively). Specifically, the Charness et al. (2012) considered a GLS and

regressed effort on wage, desired effort, and a nondelegation dummy variable. This nondelegation dummy was assigned a value of 1 when a firm opted against the delegation option and 0 in the control scenario. The results are reported in Table 5 below. The analysis revealed no negative impact on worker behavior due to nondelegation (compared to the control treatment), after accounting for wage effects, in both the Partner and the Stranger matching protocols. This led Charness et al. (2012: 2368) to their second conclusion:

RESULT 2: The increment in the effort level with delegation is not due to a negative reaction to nondelegation, but instead seems due to the positive effect of delegation, controlling for the wage received.

In line with Charness et al. (2012) we also use a second GLS in order to examine whether nondelegation elicits negative responses, specifically lower effort levels. The results, presented in Table 5, provide an intriguing finding: Only in the Partners treatment the results of Charness et al. (2012) can be verified. This means that nondelegation in the Partner treatment does not have a negative effect on worker behavior. However, in the Stranger matching scenario, a negative effect of nondelegation is observable.

Charness et al. (2012) argued that the reason for increased effort levels following delegation might not be solely due to positive reciprocity. Instead, it might have also been related to an increased sense of responsibility for outcomes, which might have led to more prosocial behavior on the part of employees. To test their hypothesis, Charness et al. (2012) compared effort levels in the Stranger treatment with delegation (SD) against the Dictator treatment, which also involved a stranger matching procedure but in which workers had to choose their own wage. If higher efforts were to be observed in the Stranger delegation treatment compared to the Dictator treatment, positive reciprocity could be implied as a response to delegation, since in the dictator treatment no delegation was possible, and the worker always chose both wage and effort levels.[5]

In fact, the statistical analysis revealed no significant difference between effort levels in the Stranger treatment with delegation and the Dictator treatment. Charness et al. (2012: 2370) concluded that positive reciprocity as a response to delegation can be ruled out, and formulated their third main result:

RESULT 3: The higher observed effort when the worker has been given the freedom to choose the wage is not due to positive reciprocity, but seems to reflect the increased responsibility that the worker has for the final outcome.

In addition to these three main findings, Charness et al. (2012) also analyzed the proportion of delegation by the firms over the 15 periods. The results show that delegation became more frequent over time, especially in long-term relationships (Partners treatment), where it was nearly twice as high in the last three periods compared to the first three. This trend suggested that firms learn the profitability of allowing the worker to set the wage. Interestingly, this increasing trend

[5] Interestingly, Charness et al. (2012: 2369) write that "Table 2 indicates that the worker's average efforts are 0.34 both in the Strangers treatment under delegation and in the Dictator treatment." However, Table 2 reports 0.34 as the mean effort in the Stranger treatment under delegation, but 0.36 in the Dictator treatment. These results are consistent with the data made available on the *American Economic Review*'s website.

Table 5 GLS Random effects regression on effort (Charness et al. 2012 and replication study)[c]

	Nondelegation partners versus control partners		Nondelegation strangers versus control strangers	
	(1)	(2)	(3)	(4)
Wage				
Original	0.006*** (0.000)		0.001*** (0.0002)	
Replication	0.002*** (0.000)		0.0027*** (0.000)	
Desired effort				
Original	0.118** (0.038)		0.040** (0.018)	
Replication	0.328*** (0.062)		0.087 (0.0671)	
Nondelegation				
Original	– 0.024 (0.036)	– 0.024 (0.061)	– 0.013 (0.017)	– 0.018 (0.018)
Replication	– 0.0310 (0.0228)	– 0.079 (0.031)	– 0.115** (0.0564)	– 0.100** (0.046)
Heterogeneity				
Original	Yes	No	Yes	No
Replication	Yes	No	No	No
Constant				
Original	– 0.069** (0.029)	0.458*** (0.038)	– 0.087 (0.057)	0.166*** (0.013)
Replication	– 0.0076 (0.183)	0.328* (0.182)	0.177 (0.225)	0.4948*** (0.1676)
R^2				
Original	0.582	0.002	0.103	0.005
Replication	0.382	0.035	0.159	0.080
N				
Original	611	611	662	662
Replication	688	688	700	700

[c]Notes: Robust standard errors are in parentheses. Delegation takes the value 1 if and only if the firm delegated the wage decision. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Also the Big Five Inventory consisting of the traits, extraversion, agreeableness, conscientiousness, neuroticism and openness was used to control for heterogeneity with each characteristic being computed as an average of two questions (Rammstedt and John 2007)

was reported as significant only in the Partners case, indicating a stronger learning effect in long-term relationships. In contrast, in short-term interactions (Strangers treatment), although there was a slight increase in delegation, it was not statistically significant, possibly reflecting the less apparent benefits of delegation in varied worker scenarios.

Table 6 Replication success for main findings

Finding	Method used	Replication sucess
RESULT 1		
Wage delegation increases worker performance	Mean comparison: Effort PND vs PD and SND vs SD (Wilcoxon matched-pairs signed-rank tests)	Yes
Wage delegation increases earnings	Mean comparison: total, firm, and worker earnings PND vs PD and SND vs SD (Wilcoxon matched-pairs signed-rank tests)	Total earnings: Yes Firm earnings: Partially (only in Stranger treatment) Worker Earnings: Yes
The wage delegation effect exists even when controlling for wage effects	GLS Random effects regression Partner vs Stranger (Explanatory variables: Wage, Desired effort, and a Delegation dummy)	Yes (Stranger) No (Partner)
RESULT 2		
There is no negative reaction to "intentional nondelegation"	GLS Random effects regression PND vs Partner Control SND vs Stranger Control (Explanatory variables: Wage, Desired Effort, and a Nondelegation dummy)	Yes (Partner) No (Stranger)
RESULT 3		
Increased effort is not due to positive reciprocity, but seems to reflect the increased responsibility of workers	Mean comparison: Effort SD vs Dictator (Wilcoxon-Mann–Whitney-Test)	No

In order to address the question raised by Charness et al. (2012) regarding whether higher worker effort in delegation cases is due to positive reciprocity, we compare the stranger and dictator treatments from our replication study. Table 3 clearly shows that the effort level in the Stranger treatment under delegation (0.63) is higher than in the Dictator treatment (0.51). This difference is statistically significant ($Z = 3.318$, $p = 0.001$, two-tailed test). Thus, the results indicate that unlike in Charness et al. (2012), positive reciprocity might indeed be a factor influencing worker behavior in terms of their effort under delegation.

Considering how firms chose delegation over the course of 10 periods, it becomes apparent that, in contrast to the findings of Charness et al. (2012), the replication study does not show a significant increase in delegation over time. A Spearman test indicates no significant upward trend in both the Partners ($\rho = -0.044$, $p = 0.371$) and Strangers ($\rho = 0.007$, $p = 0.852$) scenarios.

4 Discussion & conclusion

As summarized in Table 6, we use the main results of Charness et al. (2012) to structure our discussion of the degree to which we were able to successfully replicate the original study.

RESULT 1: Delegating the wage decision enhances worker performance and increases the earnings of both firms and workers relative with the case where firms do not delegate.

We were able to largely confirm the first result of Charness et al. (2012), showing that delegation leads to increased worker effort in both the Strangers and Partners treatments. Consistent with Charness et al. (2012), our GLS regression analysis validated the impact of delegation on effort levels in the Stranger treatment. However, this effect was not observed in the Partner treatment, which is unexpected given that firms and workers could more easily engage in strategic decision making: workers display higher effort levels in response to delegation to increase the likelihood of wage delegation in the next round.

Regarding earnings, we were able to confirm all but one of Charness et al. (2012)'s findings. In the case of the Partner treatment, the replication study's results did not show a significant difference in the average earnings of firms between cases of delegation and nondelegation.

RESULT 2: The increment in the effort level with delegation is not due to a negative reaction to nondelegation, but instead seems due to the positive effect of delegation, controlling for the wage receive.

Our GLS regression revealed that in the Partners treatment we were able to replicate the results of the original study. That is, there was no negative worker response to intentional nondelegation in this case. In contrast, we did observe a negative effect of nondelegation in the Strangers treatment, although this effect was very small compared to the influence of wage and desired effort.

RESULT 3: The higher observed effort when the worker has been given the freedom to choose the wage is not due to positive reciprocity, but seems to reflect the increased responsibility that the worker has for the final outcome.

We failed to replicate this result. When comparing effort levels across the dictator and the stranger delegation treatments, we did find significant differences which Charness et al. (2012) did not observe. So, in contrast to the original study, positive reciprocity may indeed have been an influencing factor in our sample.

Taken together, the above described results prompt the question why some of our results are inconsistent with those of the original study. It is particularly striking that in the original study the effect of delegation on effort is so much stronger *in the Partners treatment* (see Table 3 and Figs. 1, 2, 3, 4). On one interpretation, either our result is a type II error (false negative), or the result of the original study is a type I error (false positive). To further evaluate the validity of this argument, additional replication studies would be necessary.

On an alternative interpretation, the observed inconsistencies are due to the increased social distance between participants which our experimental design caused and which may have prevented a sense of togetherness among workers and employers. While wage disparities remain comparable to those in the original study, the differences in effort levels are substantially smaller in our study. This suggests that with the social distance induced by our online setting, workers were inclined to increase their own compensation when they could set wage levels by themselves, yet they did not strategically choose higher effort levels to the same degree as in the original study. Employers, in turn, could not establish trust and did not increase the rate of delegation over time. We hence did not observe a significant trend in the practice of delegation, which could have served as an indicator of the development of mutual trust. Given these tendencies, wage delegation did not increase firm earnings in our setting. These results contrast with Charness et al. (2012: 2369), who suggest that wage delegation may cause an enhanced sense of responsibility among workers, leading to a kind of virtuous circle: delegation leads to higher efforts and increased earnings for both workers and firms. Our results, on the other hand, rather support the notion that formal controls become more important as the social distance typical of remote work increases (Schmelz and Ziegelmeyer 2020).

As a final interpretative effort, we consider whether the observed inconsistencies could be attributed to differences in sample composition. Notably, our study diverges from the original study by recruiting participants from the UK via Prolific, while the participants in Charness et al. (2012) were students in Spain. Expanding the participant pool to include a general population via an online participation platform offers the advantage of enhanced external validity. Also university-based participation pools have been characterized as "WEIRD" – Western, Educated, Industrialized, Rich, and Democratic (Henrich et al. 2010). On the other hand the educational qualifications between our sample and that of the original study are remarkably similar, with over 70% of our participants possessing a tertiary education degree. Also, previous experimental research indicates that the behavior observed in student samples in economic experiments is broadly in line with that exhibited by other demographic cohorts (Exadaktylos et al. 2013). Thus, the variation in the sample composition may be considered to only have a marginal effect relative to the impact of social proximity.

Our findings provide important avenues for further research. It would be interesting to conduct the experimental procedure in both an online environment and a physical laboratory setting, concurrently procuring analogous samples from an identical population (Hergueux and Jacquemet 2015). In the online experiment, the potential for high attrition rates could possibly be mitigated by conducting the experiment simultaneous for all participants, thereby reducing variations in participant engagement and ensuring a consistent application of experimental conditions (Mason and Suri 2012). Moreover, further qualitative research could facilitate a more profound understanding of the underlying reasons behind the observed phenomenon of employers delegating less and employees demonstrating less effort.

In conclusion, wage delegation seemingly requires social proximity to be effective. These differences warrant further investigation in subsequent experiments to determine why exactly remote work seems to complicate trust-based controls.

Author contributions This paper is co-authored by Hendrik Niehoff (HN) and Philipp Schreck (PS). All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by HN and PS. The first draft of the manuscript was written by HN and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. We acknowledge financial support by the Friede Springer Stiftung.

Data availability The data that support the findings of this study are available at openICPSR: https://doi.org/https://doi.org/10.3886/E209206V1.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Albers S (2014) Preventing unethical publication behavior of quantitative empirical research by changing editorial policies. J Bus Econ 84:1151–1165. https://doi.org/10.1007/s11573-013-0675-3

Arechar AA, Gächter S, Molleman L (2018) Conducting interactive experiments online. Exp Econ 21:99–131. https://doi.org/10.1007/s10683-017-9527-2

Bloom N, Barrero JM, Davis S, et al (2023) Survey: Remote Work Isn't Going Away–And Executives Know It. Harvard Business Review. https://hbr.org/2023/08/survey-remote-work-isnt-going-away-and-executives-know-it

Brück C, Knauer T, Meier H, Schwering A (2021) Self-set salaries and creativity. J Bus Econ 91:91–121. https://doi.org/10.1007/s11573-020-00985-z
Brüggen A, Maas VS, Van Den Abbeele A (2021) Replication in experimental management accounting research - an editorial. J Manag Control 32:155–160. https://doi.org/10.1007/s00187-021-00324-5
Burbano VC, Chiles B (2022) Mitigating gig and remote worker misconduct: evidence from a real effort experiment. Organ Sci 33:1273–1299. https://doi.org/10.1287/orsc.2021.1488
Camerer CF, Dreber A, Forsell E et al (2016) Evaluating replicability of laboratory experiments in economics. Science 351:1433–1436. https://doi.org/10.1126/science.aaf0918
Chang A, Li P (2022) Is economics research replicable? sixty published papers from thirteen journals say "often not." CFR 11:185–206. https://doi.org/10.1561/104.00000053
Charness G, Haruvy E, Sonsino D (2007) Social distance and reciprocity: an Internet experiment. J Econ Behav Organ 63:88–103. https://doi.org/10.1016/j.jebo.2005.04.021
Charness G, Cobo-Reyes R, Jiménez N et al (2012) The hidden advantage of delegation: pareto improvements in a gift exchange game. Am Econ Rev 102:2358–2379. https://doi.org/10.1257/aer.102.5.2358
Charness G, Cobo-Reyes R, Lacomba JA et al (2016) Social comparisons in wage delegation: experimental evidence. Exp Econ 19:433–459. https://doi.org/10.1007/s10683-015-9448-x
Cramton CD (2001) The mutual knowledge problem and its consequences for dispersed collaboration. Organ Sci 12:346–371. https://doi.org/10.1287/orsc.12.3.346.10098
Cramton CD, Webber SS (2005) Relationships among geographic dispersion, team processes, and effectiveness in software development work teams. J Bus Res 58:758–765. https://doi.org/10.1016/j.jbusres.2003.10.006
Ditzen J, Elhorst JP (2022) Introducing the Replication Studies section. Spat Econ Anal 17:7–9. https://doi.org/10.1080/17421772.2022.2018169
Duvendack M, Palmer-Jones R, Reed WR (2017) What is meant by "replication" and why does it encounter resistance in economics? Am Econ Rev 107:46–51. https://doi.org/10.1257/aer.p20171031
Exadaktylos F, Espín AM, Brañas-Garza P (2013) Experimental subjects are not different. Sci Rep 3:1213. https://doi.org/10.1038/srep01213
Faillo M, Piovanelli C (2017) Wage delegation and intrinsic motivation: an experimental study. CEEL Working Papers
Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191. https://doi.org/10.3758/BF03193146
Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10:171–178. https://doi.org/10.1007/s10683-006-9159-4
Franke J, Gurtoviy R, Mertins V (2016) Workers' participation in wage setting: a gift-exchange experiment. J Econ Psychol 56:151–162. https://doi.org/10.1016/j.joep.2016.07.002
Giamattei M, Yahosseini KS, Gächter S, Molleman L (2020) LIONESS Lab: a free web-based platform for conducting interactive experiments online. J Econ Sci Assoc 6:95–111. https://doi.org/10.1007/s40881-020-00087-0
Hamermesh DS (2007) Viewpoint: Replication in economics. Can J Econ/revue Can D'économique 40:715–733. https://doi.org/10.1111/j.1365-2966.2007.00428.x
Harrison JS, Freeman RE (2004) Is organizational democracy worth the effort? Acad Manag Exec 1993–2005(18):49–53
Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behav Brain Sci 33:61–83. https://doi.org/10.1017/S0140525X0999152X
Hensel PG (2019) Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015. Eur Manag J 37:45–57. https://doi.org/10.1016/j.emj.2018.03.004
Hergueux J, Jacquemet N (2015) Social preferences in the online laboratory: a randomized experiment. Exp Econ 18:251–283. https://doi.org/10.1007/s10683-014-9400-5
Jeworrek S, Mertins V (2019) Wage delegation in the field. J Econ & Manage Strategy 28:656–669. https://doi.org/10.1111/jems.12313
Köhler T, Cortina JM (2021) Play it again, Sam! an analysis of constructive replication in the organizational sciences. J Manag 47:488–518. https://doi.org/10.1177/0149206319843985
Maniadis Z, Tufano F, List JA (2015) How to make experimental economics research more reproducible: lessons from other disciplines and a new proposal. Replication in experimental economics. Emerald Group Publishing Limited, pp 215–230

Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. Behav Res 44:1–23. https://doi.org/10.3758/s13428-011-0124-6

Maynard MT, Gilson LL, Mathieu JE (2012) Empowerment—Fad or fab? A multilevel review of the past two decades of research. J Manag 38:1231–1281. https://doi.org/10.1177/0149206312438773

Mellizo P, Carpenter J, Matthews PH (2014) Workplace democracy in the lab: workplace democracy in the lab. Ind Relat J 45:313–328. https://doi.org/10.1111/irj.12054

Mueller-Langer F, Fecher B, Harhoff D, Wagner GG (2019) Replication studies in economics—How many and which papers are chosen for replication, and why? Res Policy 48:62–83. https://doi.org/10.1016/j.respol.2018.07.019

Open Science Collaboration (2015) Estimating the reproducibility of psychological science. Science 349:aac4716. https://doi.org/10.1126/science.aac4716

Palan S, Schitter C (2018) Prolific.ac—A subject pool for online experiments. J Behav Exp Financ 17:22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer E, Brandimarte L, Samat S, Acquisti A (2017) Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 70:153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Rammstedt B, John OP (2007) Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. J Res Pers 41:203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Schmelz K, Ziegelmeyer A (2020) Reactions to (the absence of) control and workplace arrangements: experimental evidence from the internet and the laboratory. Exp Econ 23:933–960. https://doi.org/10.1007/s10683-020-09666-8

Semler R (2007) Out of this world: doing things the Semco way. Glob Bus Org Exc 26:13–21. https://doi.org/10.1002/joe.20161

Tuna C (2008) Pay, your own way: firm lets workers pick salary. Wall Street J 7:86

Yin Y, Wang Y, Lu Y (2019) Antecedents and outcomes of employee empowerment practices: a theoretical extension with empirical evidence. Hum Resour Manag J 29:564–584. https://doi.org/10.1111/1748-8583.12243