



In absence of money: a field experiment on volunteer work motivation

Vanessa Mertins¹ · Christian Walter¹

Received: 4 July 2019 / Revised: 18 September 2020 / Accepted: 28 September 2020 /
Published online: 4 November 2020
© The Author(s) 2020

Abstract

Although volunteers are a critical resource for non-profit organizations, little is known about how best to motivate them to work. A non-profit organization asked episodic volunteers to produce handmade greeting cards to sell at a fundraising event. By running a natural field experiment, we study the effect of motivating these volunteers through (a) the opportunity to vote on how the money that was raised would be spent and (b) the prospect of individual performance feedback. We find an economically and statistically significant positive effect of both tools on the quantity of work done, while the quality is mostly unaffected. Moreover, we observe significant gender differences in responsiveness to the treatments. While the prospect for feedback is more motivating to men, women respond more strongly to the opportunity to decide how the money would be spent. Empowerment seems to be a simple way to increase engagement for people with low enjoyment.

Keywords Autonomy · Feedback · Field experiment · Motivation · Volunteering

JEL Classification C93 · D64 · L30 · M54

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10683-020-09686-4>) contains supplementary material, which is available to authorized users.

✉ Christian Walter
christian.walter@uni-vechta.de
Vanessa Mertins
vanessa.mertins@uni-vechta.de

¹ Faculty of Educational and Social Sciences, Social Services Management, University of Vechta, Neuer Markt 32, 49377 Vechta, Germany

1 Introduction

Each year, millions of volunteers devote billions of hours to helping others.¹ While their motives for donating their time having been extensively studied,² the question of how to motivate them has not. It is important for researchers and non-profit managers to understand under what conditions volunteers increase their efforts in coping with the tasks that are allocated to them, especially in those increasingly-common settings where, on some single occasion for some short-term project, managers need to motivate large numbers of unpaid workers whom they do not know personally and who differ greatly in their motives, skills, and commitment towards the organization's mission. In such settings where financial incentives do not apply and only the conditions can raise or lower volunteers' intrinsic motivation, we study the causal effects of (a) the opportunity to vote on what the money raised will be spent on and (b) the prospect of individual performance feedback on volunteers' quantity of work.

Volunteering in projects with a limited duration on a sporadic basis is the choice of many individuals with restrictions such as increased personal time constraints, increased mobility, and irregular job obligations. Accordingly, volunteers are increasingly switching from long-term, regular, and frequent volunteer services to short-term or even one-time activities (Hustinx et al. 2008). These so-called "episodic volunteers" (Culp and Nolan 2000; Merrill 2006) are becoming a critical resource for institutions relying on external unsalaried support (Putnam 1995; Eckstein 2001; Hustinx 2001; Hustinx and Lammertyn 2003; Handy et al. 2006; Beder and Fast 2008; Smith et al. 2010; Wilson 2012). Common examples of activities they participate in include charity runs to raise funds for education, research, or support services; community events; and events to care for marginalized groups (Harrison 1995).

For such projects, the critical question is not *how to motivate people to volunteer*. Recruiting episodic volunteers is a rather easy task, since many people are intrinsically motivated to help when the commitments of time and energy are low and the benefits for those in needs are great. Accordingly, episodic volunteering events often attract hundreds of helpers (e.g. to prepare meals for patients; see Hustinx et al. 2008) or even thousands of helpers (e.g. during a crisis such as an earthquake or

¹ In the US, 62.8 million people (25.3% of the population) volunteered 7.9 billion hours (Bureau of Labor Statistics 2016), in Canada, 12.7 million people (44%) volunteered 1.9 billion hours (Turcotte 2015) and in Australia, 5.8 million people (31%) participated 743 million hours to voluntary work (Australian Bureau of Statistics 2015) in 2014.

² In their seminal paper, Clary et al. (1998) outline the following six motives for charitable giving (including time donations): 1. Values, 2. Understanding, 3. Enhancement, 4. Career, 5. Social and 6. Protective (ego defensive). Likewise to these motives, Bekkers and Wiepking (2011) find, in a literature review, eight mechanisms that drive charitable giving (including time donations): 1. Awareness of Need, 2. Solicitation, 3. Costs and Benefits, 4. Altruism, 5. Reputation, 6. Psychological Benefits, 7. Values, 8. Efficacy. Economists typically refer to pure or impure altruism. Pure altruism confers feeling a 'warm-glow' by doing "good" (Andreoni 1990, 2006; Burns et al. 2006; Brown et al. 2019). Impure altruism refers to image or reputational concerns (Ariely et al. 2009; Carpenter and Myers 2010), career concerns (Ziemek 2006; Al-Ubaydli and Lee 2011; Heinz and Schumacher 2017) or material incentives (Carpenter and Myers 2010) as potential motives for individuals to volunteer.

hurricane³ (Cnaan and Handy 2005)). Instead, the key question is *how to motivate volunteers effectively*. This question is under-researched despite its enormous practical implications (see the recent meta-analysis by Englert and Helmig 2018).

Causal evidence for ways to increase volunteers' performance is especially scarce. This is surprising, given the huge number of experimental studies in the lab and in the field studying the effects of financial and non-financial incentives on the performance of paid workers.⁴ However, evidence from those studies is not easily transferable to unpaid volunteers. While the gold standard for performance increases among standard workforces is still money (see e.g. Levitt and Neckermann 2014), monetary incentives seem to crowd out intrinsic motivation among volunteers.⁵ In a seminal experimental paper, Gneezy and Rustichini (2000) incentivized volunteers with small cash rewards. It became evident that rewarded volunteers worked fewer hours than volunteers who received no compensation. Similarly, Conrads et al. (2016) found in a field experiment that offering small monetary rewards does not help or even becomes detrimental compared to a pure volunteering treatment or the opportunity to get a certificate. On the other hand, in line with neoclassical reasoning, when volunteers in the same experiment were offered larger monetary rewards, hours they offered to work were high. So in the absence of budget constraints, money may help. Many nonprofit organizations, however, simply cannot afford to compensate their volunteers adequately. The question, then, is whether any non-financial tools are feasible. In general, non-monetary incentives such as worker recognition or adding meaning to mundane tasks can have a strong motivational impact on work performance of paid workers (e.g. Kosfeld and Neckermann 2011; Ashraf et al. 2014; Kosfeld et al. 2017), but again, such means of motivation are not easily transferable to volunteers, especially not to episodic ones. However, it isn't impossible to motivate volunteers with non-monetary incentives. Bradler et al. (2016) conducted a natural experiment in which a subset of students hired for a one-time data-entry job unexpectedly received thank-you cards during task performance. The output of the students in the subset increased significantly following the expression of gratitude. At the same time, several non-monetary incentives that have recently been studied for paid workers would be hard to transfer to a volunteer context. These include the implementation of competition among workers (Irlenbusch and Ruchala 2008; Cardinaels et al. (2018), as well as non-material benefits such as career opportunities and preferred parking (Barbara et al. 2017).

The volunteering context seems to be simply too different to apply ordinary motivational factors from normal working life. This point is underscored by the few studies that have so far attempted to investigate a causal relationship between the factors of the work situation that can be manipulated and the performance of volunteers. Al-Ubaydli

³ After Hurricane Harvey devastated wide areas of Texas in 2017, tens of thousands of US citizens helped spontaneously with their bare hands, with food supplies and with their machines and tools.

⁴ For a survey of field experiments involving firm-worker relations, see e.g. Levitt and Neckermann (2014) or List and Rasul (2010).

⁵ The crowding out of intrinsic motivation by monetary incentives has been demonstrated in several studies, e.g., Bénabou and Tirole (2006), Mellström and Johannesson (2008), Ariely et al. (2009), Carpenter and Myers (2010), or Meer (2014), Bowles and Polania-Reyes (2012) and Conrads et al. (2016).

and Lee (2011) provided volunteers with tailored communication via newspaper articles according to their individual predominant motive for their volunteer engagement. Compared to a control group, the results show only a small positive effect on volunteered hours, and only for those volunteers who are mostly motivated by career concerns. Conrads et al. (2016) find potential detrimental effects of seemingly useful rewards, such as waiving a fee for student volunteers helping at a conference, which massively decreased the number of work hours offered.

When thinking about how to inspire people to perform best, it is crucial to create an atmosphere within people will increase their intrinsic motivation themselves. Such intrinsic, autonomous motivation is especially important in voluntary work, which is defined as performing an activity of one's own free will, without formal obligations or remuneration (Bidee et al. 2013). Volunteers usually contribute to the greater good as long as they feel they are making a meaningful contribution. We build upon the concept of autonomous (also called self-determined) motivation, as it refers to engaging in an activity of one's free will or with a sense of choice (Ryan and Deci 2000; Grouzet et al. 2004). The underlying self-determination theory (Deci and Ryan 1985; Ryan and Deci 2000) is one of the leading theories of human motivation. It posits that support in volunteers' autonomy is the way to create the conditions within which other people will motivate themselves.

In this paper, we therefore experimentally test the impact of two conditions intended to increase autonomous motivation on the performance of unpaid workers. Our partner organization supports children with cardiac problems, raising funds through charitable donations and through the sale of goods. For an upcoming summer fair, the organization invited volunteers to a one-time social project about 1.5 h long. In total, 190 volunteers (mostly college students), showed up and agreed to produce handmade greeting cards by coloring mandala paints. We used this setting to test under which conditions different types of volunteers in the absence of pay would increase their performance. In the *VOTE* treatment, we allowed volunteers to vote for which purpose the earned money should be spent, while in the *FEEDBACK* treatment we announced that they would receive individual performance feedback after the work was done.

We observe a statistically and economically significant performance increase between both treatments and the control group. Volunteers in *FEEDBACK* colored 36.5% more cards and those in *VOTE* colored 44.6% more cards, while the cards' quality does not differ significantly between treatments. Male volunteers respond more positively to the feedback announcement, while females get strongly motivated by the opportunity to vote for the preferred purpose. Both treatments work especially well among volunteers with low task enjoyment, a subgroup that has low intrinsic motivation. These findings support the idea that in the absence of monetary incentives, simple cost-effective tools—easy to adapt for diverse individuals—can be designed to boost the performance of volunteers.

2 Conceptual framework

Increasing the performance of volunteers is not a simple task, especially because incentives that seem reasonable at first glance (such as pay for performance) have been shown to backfire. Indeed, extrinsic rewards seem to crowd out intrinsic motivation (Gneezy and Rustichini 2000), at least for low monetary incentives and relatively weak extrinsic motivators (Conrads et al. 2016). The vast majority of people volunteer not for monetary reasons but many for a so-called “warm glow,” a positive feeling (Andreoni 1989; Ellingsen and Johannesson 2009; Bauer et al. 2013, Brown et al. 2019). Volunteers “display a stronger desire to donate time and effort than to donate money” (Brown et al. 2019, p. 33). Mentioning extrinsic rewards seems to reduce intrinsic motivation among volunteers (Conrads et al. 2016). A crucial factor that explains the variability in intrinsic motivation is the perceived meaningfulness of a task. Experimental evidence from the lab (Ariely et al. 2008) and the field (Chandler and Kapelner 2013; Chadi et al. 2017; Bäker and Mechtel 2018) suggests that people perform better when they feel that what they do is meaningful. Since typical voluntary tasks—such as stuffing letters, sorting donated clothes, or preparing standardized food packages—are often simple and repetitive (Smith et al. 2010), easy to understand by mostly inexperienced volunteers (Hyde et al. 2014), and independent of volunteers’ cognitive or crafting skills (Hustinx et al. 2008), the *task* itself usually provides no source of meaningfulness. Therefore, in volunteer settings, it is even more important that the entire work environment is designed to be as appealing as possible.

Psychologists emphasize that a social environment can stimulate, hinder, or block the positive features of human nature. The conceptual framework underlying this observation is the self-determination theory of Deci and Ryan (1985) and Ryan and Deci (2000), which defines three innate psychological needs for motivation: (1) competence (the feeling of being able to work efficiently and achieve the desired results) (2) autonomy (being able to choose one’s actions from a range of possibilities) and (3) relatedness (feeling connected to other human beings). In the seminal papers, Gagné (2003) and Gagne and Deci (2005) demonstrate that self-determination theory can be applied to the domain of prosocial motivation, and suggest encouraging social structures that fulfill basic psychological needs. People who feel competent, and related to their peers may be more likely to be motivated to engage in more prosocial behaviors. While survey evidence (e.g. Haivas et al. 2012; Bidee et al. 2013) support this notion, there is—to the best of our knowledge—no experimental study so far analyzing the impact of autonomy-supporting contexts on the actual performance of volunteers.

We designed two treatments that were intended to give people choice and encouragement for personal initiative and also support their feelings of relatedness. In the FEEDBACK treatment, we announced that volunteers would receive individual performance feedback after the work was done. This was expected to satisfy individuals’ basic psychological need of relatedness and competence in that people can better learn how they contribute to the greater good. As self-determination theory suggests, being involved in activities that benefit society

and feeling related to other human beings, including to a group or the organization, can enhance intrinsic motivation (Ryan and Deci 2000). Feeling connected and engaged in the activity helps to strengthen volunteers' relationship with the nonprofit organization and its members. Further, the prospect of feedback is likely to increase the feeling to perform adequately. The economic literature on the provision of feedback mostly suggests positive effects, such as on educational outcomes (Azmat and Iriberry 2010), student performance (Bandiera et al. 2015), work performance (e.g. Falk and Ichino 2006; Gerhards and Siemer 2015), and task motivation (Banuri et al. 2018). A few studies (e.g. Eriksson et al. 2009, Azmat et al. 2019) report detrimental outcomes.

In the *VOTE* treatment, we allow volunteers to vote for which purpose the earned money should be spent. This treatment supports both autonomy (people have a say in the purpose of the money to be raised) and relatedness (people may contribute to the greater good in the way they like). By allowing volunteers to participate in the decision-making process, presumably their perceived freedom and choice was increasing. Besides, this treatment allows people to feel that their contribution matters for the greater good; they experience what they have contributed to the good cause by deciding the purpose of their work. While the causal effect of such a strategy has not been analyzed, there is some related evidence on the impact of participation in the decision-making process on workers' performance. Charness et al. (2012) and Jeworrek and Mertins (2019a) varied workers' participation in the determination of wages. Allowing workers to self-determine their wages yielded performance-enhancing effects that were smaller in the field (Jeworrek and Mertins 2019a) than in the lab (Charness et al. 2012). Related studies also show mixed effects. Franke et al. (2016), for example, find adverse incentive effects of participation that presumably were triggered by negative reciprocity. It has been, however, an open empirical question how volunteers respond to the possibility of a majority vote on what the funds generated are used for.

Following self-determination theory, and in contrast to standard economic theory, we expect both treatments, *FEEDBACK* and *VOTE*, to significantly increase volunteers' average performance (measured in number of cards produced) compared to the control group. Our rich data set makes it possible to take the heterogeneity among the volunteers (Dunn et al. 2016) into account—their varying levels of enjoyment in doing the task, their competitiveness, their solidarity with the charity. Given the literature on differences in preferences (see e.g., Niederle and Vesterlund 2007; Croson and Gneezy 2009), we expect to see that the treatments work differently for (1) male and female volunteers, (2) high and low task enjoyment volunteers, and (3) high and low competitive volunteers.

3 Experimental design

3.1 Field setting

Our charity partner, Kinderherzhilfe, is a nonprofit organization (NPO) run by the families of children with heart problems. It helps those children and their families with both monetary and non-monetary support. The organization's social events are

important for the children and their families to overcome stressful situations after medical treatments, and they allow the sharing of information. Furthermore, the organization regularly fulfills the wishes of affected children. To finance its projects, the organization raises money from donations, membership fees and commodities that it sells at flea markets and regional celebrations. For a fundraising event at the upcoming summer fair, we partnered with the NPO and agreed to host an event where a large group of volunteers would produce hand-colored greeting cards. The greeting cards had a mandala on the front page, a blank space in the middle for individual texts and the logo of the organization on the back of the card. The mandala was blank and had to be colored by hand, giving it an individual touch in order to raise more money through a higher selling price. We announced that there would be an event at which volunteers would do a one-time task for a good cause. To attract every potential volunteer, including the subgroup of supporters who did not enjoy coloring, we did not mention the specific task in the announcement. The upcoming event was announced at several lectures and was promoted as a special occasion to support a local charity with an urgent one-time task. Since the university is widely renowned for its focus on social studies subjects—such as social work, gerontology, and social services management—many charities regularly host events on the campus or recruit volunteers among a large number of socially-engaged students.

3.2 Work task

We took the opportunity that the charity planned to sell greeting cards at the upcoming summer fair to raise funds. The work task was prepared in cooperation with the charity to fit into the nature of the episodic event: it had to be temporary, feasible, and related to children. Also, the task had to be performed regardless of expertise or skills, since we expected volunteers with different backgrounds and sociodemographic characteristics to take part. While hand-colored cards are often used by NPOs to foster charitable giving, the task of preparing them had never been used by experimental economists. The task allowed us not only to measure the quantity of work done but also to control for the quality of the work.

The volunteering event attracted students from various fields. It had a friendly, informal atmosphere. The volunteers could leave whenever they wanted and knew that they would not get paid for the work task. Furthermore, it was made clear that it was a one-time event. After the volunteers entered the room, we assigned them a random treatment. The seating instructions were provided naturally, politely, and informally. By explaining that there were only three sets of premium pastel crayons that were essential for coloring pretty cards, we were able to allocate them into three treatment groups of similar size.⁶

⁶ Spillover effects between treatment groups can potentially harm the explanatory power of experimental studies. We took great care to prevent them through various measures. These included strict spatial separation between treatment groups based on the fact that only three of the expensive pastel crayon sets were available, adequate space between individuals as well as between treatment groups, and very precise instructions which prevented the volunteers from having to ask questions in the plenum about their individual treatment. Furthermore, we took great care that, at first glance, the instructions looked identical, so the volunteers stayed unaware during the experiment that there were different treatment groups.

In the beginning, a member of the charity welcomed the participants and gave a short introduction. During the coloring event, each volunteer got two handouts and one questionnaire. The first sheet, *Handout 1* (see Fig. 7 in Appendix), introduced the event, promised anonymity provided an individual supporter number to each participant, and included a small survey that asked about the respondents' previous knowledge about the charity (1 item) and willingness to support it (5 items). We gathered the information in the survey to ensure that no unobserved factors would bias the potential treatment effects. In the next step, the member of the charity promoted the organization's mission and goals, showing some pictures from previous activities and campaigns. Emphasizing that volunteers are supporting a good cause has recently been shown to be a significant positive determinant of their motivation to perform well (Carpenter and Gong 2016; Jeworrek and Mertins 2019b).

3.3 Treatments

When the oral introduction ended, *Handout 2* was distributed. It summarized the purpose of the event and introduced the task as well as the information that 100% of the money raised would be spent on charitable projects (see Fig. 8 in Appendix). It also announced that the charity needs the help of the present volunteers to color as many and as beautiful cards as possible because the NPO planned to sell them at the forthcoming summer fair using a pay-what-you-want mechanism. *Handout 2* also contained treatment texts which varied in one sentence only.

In the CONTROL treatment, no further information was given. In the FEEDBACK treatment, we announced that the individual selling results—the amount of money that each person's cards sold for would be published on a webpage some days later anonymously. We wanted to test whether the promise of feedback would increase volunteers' performance. Since the results were anonymous and therefore private, the setting minimized image concerns and image motivation (Ariely et al. 2009; Goette et al. 2010); in other words, coloring more mandalas in order to be more liked or respected by others should play no decisive role. In the VOTE treatment, volunteers were given the right to vote for the purpose on which the money they raised would be spent. People could choose between the purposes by ticking the respective box: organizing a barbecue, sponsoring a family holiday, printing flyers, fulfilling children's wishes, or buying toys for the cardiology ward of the local hospital.⁷ In Table 1 we show the different treatment texts in direct comparison.

Volunteers had plenty of time to read *Handout 2*, which included the treatment texts, while student helpers distributed the blank greeting cards. After the pastel crayons were handed out, the coloring session started. Whenever volunteers finished drawing a card, they could take another one from a pile in front of them. The drawing session went smoothly and uninterrupted. After forty-five minutes, we thanked all the participants. Subsequently, a group of students asked the volunteers

⁷ To keep the number of questionnaire answers constant across treatments and to provide each individual with the same information (except treatment texts), individuals in the other treatments could also read the different purposes of how to spend the money.

Table 1 Treatments

Control	Feedback	Vote
<i>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can.[...]</i>	<i>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can.[...]</i>	<i>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can.[...]</i>
	<i>After the event, there will be an anonymous list saying how much each one of you earned for the good cause</i>	<i>As a supporter, you are allowed to vote for which charitable purpose the collected money will be spent</i>

to complete a questionnaire for an ongoing research seminar.⁸ This questionnaire included items on individuals' demographics (age, gender), task enjoyment, and competitiveness (see Fig. 9 in Appendix for the exact wording).⁹ At the end, we collected all materials, sorted by individual participant.

3.4 Quality

The cards' quality was likely to differ between volunteers: They could use up to 80 different colors. Some volunteers used a wide color spectrum, while others used just one crayon per card. Besides, the color intensity could be varied through pressure, and the colors could be mixed with a tiny drop of water or oil for some special effects.

To assess the quality of the cards, we used two different approaches. First, we asked three research assistants not involved in the conduct of the experiment to independently assess the subjective quality of each card by rating the subjective overall impression on a scale from 1 ("moderate") to 5 ("very handsome"). The average rating from the three assessments gave us a subjective quality evaluation for each card. Then we calculated the averages over all cards per volunteer to get one subjective quality indicator per individual.¹⁰ This variable is called *rating*. Second, we asked research assistants to assess every card based upon two characteristics: *colors*, the number of colors used; and *areas*, the number of areas colored. This resulted in two separate measures per card. Again, we calculated the averages over all cards per volunteer to get two objective quality indicators per individual. For an impression about differences in card quality, see Fig. 1.

⁸ The implementation of the field experiment was accompanied by a group of students who wrote a paper on the data from the post-experimental questionnaire. Three of them were on site and handed out the questionnaire that they had previously worked out. The students announced the questionnaire on their own with the words: "For a student research project, we will now distribute a short questionnaire and would be happy if you could fill it out." In line with the convention in field experiments in economics, we did not apply debriefing and preferred not to give any hints about the experiment.

⁹ The statements about task enjoyment were related to the Intrinsic Motivation Inventory by Ryan and Deci (2000) and asked for volunteers' enjoyment during the task. Statements about competitiveness were related to the Competitive Index by Smither and Houston (1992) (revisited by Harris and Houston 2010) to measure volunteers' competitiveness.

¹⁰ The interrater reliability is given (Cronbach's alpha = 0.73).



Note: The mean rating of the left card was a 5 (handsome). It has 5 colors and 4 colored areas: hearts, dots, stars, flowers. The middle card has a 3 rating and has 4 colors and 5 colored areas. The right card was rated 1 (moderate) and had 1 color and 4 colored areas.

Fig. 1 Examples of differences in individuals' quality indicators

3.5 Selling

Four days after the coloring event, the summer fair took place. The selling booth of the charity was decorated with posters and retractable banners introducing the charity (see Fig. 2). The greeting cards were hung in random order on washing lines to ensure that customers could compare them and choose their favorite. It was announced that the greeting cards were being sold to support cardiac children and their families and that one hundred percent of the money raised would be used for children's wishes (as chosen by the majority of volunteers' votes). Buyers were allowed to pay as much or as little as they wished for the cards (pay-what-you-want). We noted not only the price, but also the *sales rank*.¹¹ From time to time, the placement of the cards hanging on the washing lines was randomly changed. The selling variables *sold cards*, *sales rank*, and *sales* (each per individual) were again calculated at the volunteer level and used as three additional quality indicators in our analysis.¹²

Two weeks later, the results were presented to all volunteers through a public posting. We showed the realized donations per individual, anonymized by showing the supporter number next to the amount raised. Further, it was revealed that the money would be spent on individual children's wishes (the outcome of the vote). Figure 3 summarizes the chronology of events in the field experiment.

¹¹ *Sales rank* is the ordering in which the cards were sold (first sold card = 336, last sold card = 1, not sold = 0).

¹² Customers' choices add some additional subjective information on the quality of the volunteers' work. *Sales rank*: Whenever a card was sold, an assistant registered the card's ID and the ranking order. Since the cards were displayed on washing lines showing all at a glance, we can assume that the more attractive cards were sold earlier. *Sales*: the price of the cards was not fixed but self-chosen by customers, using a pay-what-you-want mechanism, so prettier cards can be expected to have fetched higher prices. *Sold cards*: the absolute number of sold cards is also expected to correlate with the quality of the work done.



Notes: Both pictures display the selling booth with the greeting cards on the washing lines.

Fig. 2 Selling booth

Event	Episodic event					Afterward	4 days later	2 weeks later
Heading	Introduction	Handout 1	Handout 2	Task (45 minutes)	Questionnaire	Quality ratings	Selling	Announcement of results
Content	Welcoming	Short survey: willingness to support	Description of the task + treatment texts + vote	Painting mandalas on greeting cards	Survey: socio-demographics, task enjoyment, and competitiveness	Quality: rating, colors, and areas	Sales rank, sold cards, sales	Realized earnings, purpose

Notes: The Figure shows the chronological process of the experiment, starting with the episodic event (introduction followed by Handout 1, Handout 2, the task and the Questionnaire), afterward (quality ratings), 4 days later (selling) and 2 weeks later (announcement of results).

Fig. 3 Chronology of events

4 Results

4.1 Descriptive statistics

In total, 187 volunteers participated in the episodic volunteering event. Table 2 summarizes the descriptive statistics. Volunteers' ages ranged from 18 to 37 (mean = 22.02, SD = 2.69) and 73.6 percent were female. Most of them were majoring in social services (social work, management of social services, gerontology). Accordingly, our sample reflects not only the population of the local student body but also a composition common for episodic events (Hustinx and Lammertyn 2003; Dunn et al. 2016). The volunteers colored a total of 667 greeting cards, with individual performance differences ranging from zero to ten. The median of colored cards per volunteer was three, and the average was 3.56 cards (SD = 1.71). About half the cards (n = 324) were sold at the summer fair with an

Table 2 Descriptive statistics on volunteers' characteristics

	All (N=187) Mean (SD)	Control (N=63) Mean (SD)	Feedback (N=61) Mean (SD)	Vote (N=63) Mean (SD)	Kruskal– Wallis rank test (<i>p</i> -values)
Age	22.022 (2.69)	21.887 (0.698)	21.328 (1.491)	22.838 (0.790)	0.147
Gender	0.737 (0.441)	0.698 (0.462)	0.721 (0.452)	0.790 (0.410)	0.482
Willingness to support	3.913 (1.018)	3.904 (1.291)	4.030 (1.142)	3.809 (1.513)	0.517
Task enjoyment	4.282 (1.869)	4.640 (1.670)	4.101 (2.051)	4.087 (1.857)	0.251
Competitiveness	3.859 (1.351)	3.690 (1.405)	3.893 (1.301)	4.000 (1.345)	0.493

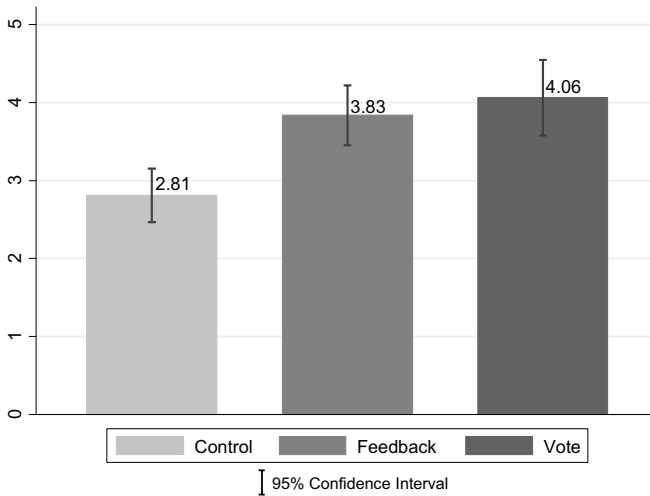
Gender is a dummy variable with 1 = female

average price of 1.13 € per card (min = 0.20 €, max = 10.00 €).¹³ From 80 percent of the volunteers, the charity sold at least one card, while the number of sold cards per volunteer ranged from zero to seven.

Most of the volunteers did not know the charity beforehand. On a Likert scale from 1 (disagree completely) to 7 (agree completely) with the statement “I’m familiar with Kinderherzhilfe,” the mean value was 1.6 (SD = 1.31) and 75% chose “1.” Although only a few volunteers had heard about the charity beforehand, many volunteers showed up to support the event. This finding is in line with former research by Cnaan and Handy (2005) and Hyde et al. (2014) who found that spontaneous interest in an opportunity to support an unknown charity is typical for episodic volunteers. This turns out to be the case in our field setting as well, since despite the low popularity of the charity, nearly 60% of volunteers chose 6 or 7 for the statement “Supporting children with cardiac problems is important to me” (mean = 5.65, SD = 1.32).

When asked about their willingness to support the charity, a majority expressed willingness to donate money (mean = 3.7, SD = 1.71) and time (mean 3.8, SD = 1.76), and most of them also said they were willing to promote the charity among friends (mean = 4.9, SD = 1.57) and strangers (mean = 3.6, SD = 1.8). We combine the item on their knowledge about the charity and the five items about their support readiness to the variable *willingness to support* (Cronbach’s alpha = 0.71). Also, we combine the three items from the *Questionnaire* asking for volunteers’

¹³ The remaining cards were given away later for promotion.



Notes: The Figure displays the average painted cards per treatment with 95 percent confidence intervals.

Fig. 4 Treatment differences in total colored cards

task enjoyment (Cronbach's $\alpha=0.88$) and four questions on volunteers' *competitiveness* (Cronbach's $\alpha=0.77$). On average, volunteers' task enjoyment was 4.28 (SD = 1.87) and competitiveness 3.86 (SD = 1.35). Regarding the random allocation into the treatments, the *p*-values of the Kruskal–Wallis rank tests for differences between the treatment groups (last column of Table 2) suggest no significant differences in volunteers' age, gender, willingness to support, task enjoyment, or competitiveness.

4.2 Treatment differences

Our main interest is the analysis of potential treatment effects on volunteers' performance. According to the conceptual framework, we test whether an increase in autonomous motivation through individual feedback provision and the opportunity of voting for the beneficial purpose leads to an increase in performance.

We find large treatment differences in the quantity of work done. In CONTROL, volunteers colored an average of 2.81 greeting cards, while volunteers in FEEDBACK and VOTE colored 3.84 and 4.06, respectively (see Fig. 4). This means that volunteers in the FEEDBACK group colored 36.5% more cards and those in VOTE 44.6% more, than did those in CONTROL. The pairwise comparisons between CONTROL and FEEDBACK as

Table 3 Main regressions

Specification	Total painted cards		
	(1)	(2)	(3)
Feedback	1.027*** (0.262)	0.969*** (0.263)	1.125*** (0.262)
Vote	1.254*** (0.300)	1.182*** (0.297)	1.300*** (0.302)
Age		-0.0117 (0.0487)	-0.0105 (0.0509)
Gender		0.478* (0.281)	0.430 (0.310)
Willingness to support			0.0373 (0.129)
Task enjoyment			0.177** (0.0748)
Competitiveness			-0.0661 (0.0862)
Constant	2.810*** (0.174)	2.771** (1.150)	2.039 (1.254)
Observations	187	185	181
R ²	0.102	0.112	0.157

The table presents ordinary least square estimates of treatment effects. In all five reported columns from (1)–(5), the dependent variable is the total amount of colored cards. The reference group in all estimations is Control. *Willingness to support* is the mean of six questions concerning someone's willingness to support the charity. *Task enjoyment* is also a mean of three task-related enjoyment statements of drawing mandalas. *Competitiveness* is the mean of four related statements. Robust standard errors are displayed in parentheses

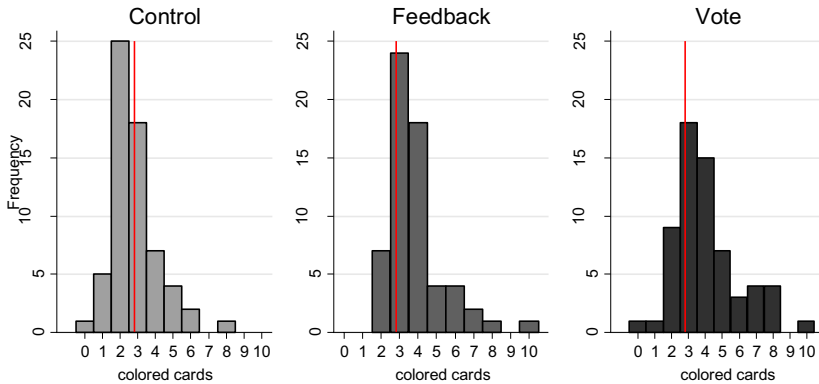
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

well as CONTROL and VOTE yield statistically and economically relevant differences, with $p < 0.0001$.¹⁴

We run three OLS regressions (Table 3) to control for different variables which were taken from *Handout 1* and the *Questionnaire*. In Column 1, we replicate the results from the non-parametric tests. In Column 2 we add two sociodemographic controls with age and gender, while Column 3 additionally adds volunteers' *willingness to support*, *task enjoyment*, and *competitiveness*.¹⁵ In any of the specifications,

¹⁴ Reported p -values are drawn from two-sided Mann–Whitney U tests unless otherwise stated.

¹⁵ Although these variables are endogenous, we find no differences amongst these between treatments, and excluding them does not alter the results, neither in terms of significance nor in terms of coefficient size.



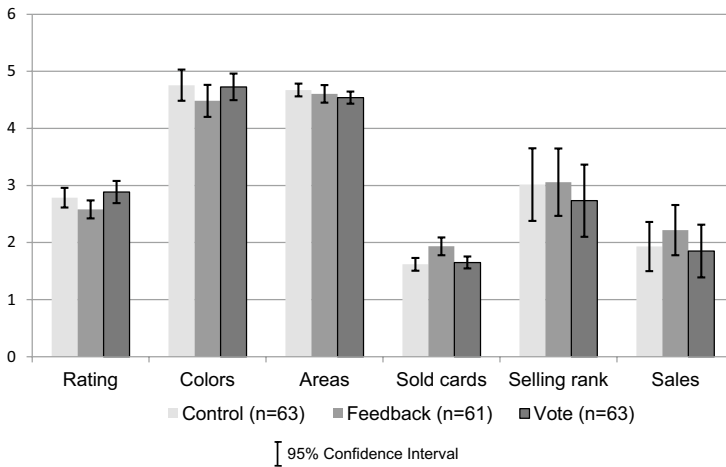
Notes: The graphic displays histograms per treatment for total colored cards. The red line displays the mean value of the control group.

Fig. 5 Histogram of colored cards by treatment

we observe highly significant treatment differences for both `FEEDBACK` and `VOTE`. The findings are completely robust when including any control variables, and the effects are sizeable and economically relevant.

While age does not affect the number of colored cards, we find that women drew on average 0.47 to 0.79 cards more than men. *Task enjoyment* is a significant predictor of volunteers' performance in terms of quantity in all specifications. This underlines the importance of task characteristics for individuals' motivation. Volunteers' *willingness to support* and their level of *competitiveness* are not predictors for their performance, probably because participants were highly homogenous in this regard. Although the quantity and quality outcomes are highly correlated (and endogenous), we find robust treatments effects on quantity when we control for the quality measures *rating*, *colors*, *areas*, *sold cards*, *sales rank* and *sales* (see Table 8 in Appendix). Thus, `FEEDBACK` and `VOTE` motivate volunteers to color more cards but do *not* crowd out beautifulness. Further, the treatments not only provided positive effects for the upper end of the performance distribution (outliers pull the effect); a further analysis reveals significant effects for all levels (see Fig. 5). As seen in the distribution of total colored cards, every volunteer in the treatment group colored on average one card more than their control group counterparts. This finding holds also true when analyzing the top performers in each treatment (fourth quartile). Top-performing volunteers in `FEEDBACK` and `VOTE` outperformed those in `CONTROL` in terms of quantity ($p=0.0251$ and $p=0.0077$) and rating ($p\text{-value}=0.0013$ and $p=0.0223$).

Last, we analyze if the treatments differ in their effectiveness. A Wald test for equal coefficients shows no significant differences among treatments regarding effectiveness. In the subsequent section, we will analyze the effectiveness of both instruments for specific subgroups.



Notes: The figure displays means on all quality indicators. The bars represent 95% confidence intervals.

Fig. 6 Treatment differences for cards quality measures

4.3 Further results

In this section, we take a closer look at the six quality measures—*rating*, *colors*, *areas*, *sold cards*, *sales rank*, *sales*—and then at the heterogeneous treatment effects. Among our volunteers, we observe large individual differences in all quality measures. Take, for example, the subjective quality measure *rating*, ranging from 1 to 5. While some volunteers reached the maximum rating of 5 with almost every card, others reached only averages of around 1.5. Similarly, the charity sold seven cards from some volunteers and zero from others. We report additional descriptive statistics in Table 9 in Appendix.

Despite the large individual heterogeneity, we do not observe reliable treatment differences regarding any of the six quality measures. Figure 6 shows the treatment comparisons with bar graphs including 95% confidence intervals for each quality indicator. It does not suggest that one treatment clearly outperforms any other in terms of quality output. As a robustness check, we run multivariate regressions on all quality measures with the standard set of control variables: *age*, *gender*, *willingness to support*, *task enjoyment*, and *competitiveness* (see Table 10 in Appendix). There is a small positive impact of *VOTE* on *sold cards* and *sales*; and *FEEDBACK* on *rating* but the results are far from conclusive.

However, once we analyze this large set of outcome variables, we should consider multiple hypothesis testing. The multiple hypothesis testing (List et al. 2019) confirms the former results. There are no noteworthy differences in quality (see Table 11 in Appendix). Thus, the sizeable increase in quantity has not crowded out the quality of produced cards.

For an adequate understanding of the following discussion of results, note the Pearson correlations between total colored cards and *rating* ($-0.2431, p=0.0009$), colored *areas* ($-0.2285, p=0.0018$), *colors* used ($-0.2864, p=0.0001$), *cards sold*

Table 4 Quality differences within and between volunteers

	Within volunteer SD	Within variation coefficient (%)	Between volunteer SD	Between variation coefficient (%)	Obs
Rating	0.377	13.5	0.646	23.2	185
Colors	0.736	15.6	1.040	22.1	185
Areas	0.221	4.7	0.513	11.1	185
Sold cards	0.494	79.4	0.268	42.3	153
Sales rank	44.496	26.3	77.189	45.6	153
Sales	0.318	21.1	1.636	67.0	153

The table presents standard deviations within and between volunteers as well as variation coefficients for the quality measures rating, colors, areas, sold cards, sales rank, and sales

(0.5284, $p=0.000$), and *sales* (0.5096, $p=0.000$). A trade-off between more cards, on the one hand, and prettier cards, on the other hand, is intuitive. Coloring prettier cards needs more patience and is, therefore, more time-consuming. On the one hand, more beautiful cards with more colors and/or more colored areas should result in fewer cards but may generate higher earnings. On the other hand, if a person colors more cards, there may be a higher chance that those cards will be bought and thereby increase the higher absolute earnings. Despite these considerations, while volunteers in both the FEEDBACK and VOTE treatment conditions colored significantly more cards, their cards were judged no less beautiful than those of the individuals in the CONTROL group.

We argue that the volunteers' coloring skills were fixed short-term and could not be increased as easily as could the effort that the volunteers exerted, especially given the tradeoff between quantity and quality. As a result, we assume that volunteers expressed their motivation through an increase in quantity rather than through an increase in quality. This argument is supported by further statistical findings. Table 4 shows standard deviations (SD) *within* volunteers (i.e. the differentiation between one person's cards) and *between* volunteers (i.e. differentiation among different volunteers). First, the intrapersonal deviation is relatively small, with a coefficient below 25%. The intrapersonal rating deviates 0.377 points and the interpersonal rating 0.646 points. We find that the mean variation coefficient of each volunteer within his or her own cards is smaller than the coefficient comparing the volunteers together. Along with the relatively small within deviations at all and the statistically equal deviations between treatments seen in Table 4, this finding supports the assumption of fixed personal drawing skills.

An alternative explanation might lie in the inefficient use of working time by volunteers in the control group. More precisely, the difference in total colored cards may not come from differences in coloring speed but rather in boredom resistance. In other words, volunteers in the control group might have taken more breaks and stopped working before the official end. However, there are a few facts that contradict this argument. First, we did not observe abandonment or unwillingness among any individual. Volunteers seem to have colored continuously and no-one left the coloring session ahead of time. Second, volunteers reported high enjoyment during the task (see Table 2). If volunteers in the CONTROL group had gotten bored drawing

mandalas, we would have seen a difference in their task enjoyment, which is not the case. From this data, our observations, and informal conversations after the event, differences in levels of enjoyment during the event can be ruled out. Third, volunteers arrived and stayed voluntarily at the event to support the good cause. It would have been unusual for volunteers to stop working during a charity event.

It is reasonable to assume that volunteers colored each mandala as beautifully as they could, and every time their personal threshold was reached, they started a new card instead of improving the former one. Therefore, the absence of quality effects is likely not a result of a lack of motivation but rather a lack of skill which ended in an optimization strategy to color as many cards as possible with a fixed quality standard. Thus, with the trade-off between more versus better cards, it is not appropriate to expect improvements in terms of quality. However, the task has its merits, above all the fact that it is a very common task in nonprofit organizations. Letting volunteers color mandalas requires little to no explanation and can be done by nearly everyone. It is easy to implement and a large number of people can perform it at once. The quantitative output is easy to measure and meaningful for both producer and charity. Further, coloring cards is not affected by the pitfall of teamwork. As quality differs largely between volunteers, but only slightly within them, it is nonetheless crucial to control for beauty.

While we found that both *FEEDBACK* and *VOTE* instruments motivate volunteers to color significantly more cards in the aggregate, we do not know yet how different types of volunteers react to the treatments. This is intriguing from a scientific and practical point of view because volunteers' motivations and characteristics are typically highly heterogeneous (Clary et al. 1998; Hyde et al. 2014),¹⁶ and thus they likely vary in their susceptibility towards the prospect of getting performance feedback or participation opportunities in the decision-making process. We focus on three different subgroups since many studies suggest gender (Croson and Gneezy 2009), task enjoyment (Ryan and Deci 2000) and competitiveness (Niederle and Vesterlund 2007) to be relevant impact factors for individual motivation.

4.3.1 Gender

Gender is an interesting variable for at least two reasons. First, imagine any episodic volunteering event where participants show up mostly unannounced and spontaneously (Cnaan and Handy 2005). A nonprofit organization faces the challenge of having scarce information about episodic volunteers and this makes categorizations in different groups even harder than for regular volunteers. Among the visible characteristics, gender is easily accessible and therefore suggests itself as an appropriate candidate. Second, there is a large literature on gender differences in preferences (see, for example, the extensive literature review by Croson and Gneezy 2009).¹⁷

¹⁶ This is particularly true for episodic volunteers whose different motivations to participate in occasional events were recently surveyed (Dunn et al. 2016).

¹⁷ The literature suggests that women are more risk-averse than (Eckel and Grossman 2003), less competitive and self-confidence (Niederle and Vesterlund 2007; Gneezy et al. 2009; Niederle and Vesterlund 2010). Further, there seem to be gender differences in pro-social behavior. Tonin and Vlassopoulos (2010), for example, find that women react to action-oriented altruism more strongly than to output-oriented altruism, while there are no differences among men.

Table 5 Heterogeneous treatment effects by gender

Specification	Male		Female	
	(1)	(2)	(3)	(4)
Feedback	2.077*** (0.525)	1.884*** (0.539)	0.591** (0.342)	0.924*** (0.291)
Vote	0.842 (0.566)	0.461 (0.486)	1.256*** (0.333)	1.502*** (0.356)
Controls		Yes		Yes
Constant	2.158*** (0.360)	3.791 (2.542)	3.091*** (0.242)	3.050* (1.720)
Observations	49	48	137	133
R ²	0.255	0.308	0.096	0.168

The table presents ordinary least squares estimates of treatment effects. In all stated columns from (1)–(4), the dependent variable is the total number of colored cards. Controls include age, willingness to support, task enjoyment, competitiveness. The reference group is CONTROL. Robust standard errors are displayed in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Therefore, it cannot be ruled out that in general, women and men react differently to the present motivational tools.

In our voluntary work situation, we observe significant gender differences in quantity ($p = 0.028$, two-sided Mann Whitney U-test), with women coloring more cards (mean = 3.7) than men (mean = 3.1). At the same time, we also observe significant gender differences in quality, at a 5 percent level, for most measures—*rating*, *sold cards*, *sales rank*, *sales*—with women outperforming men. To test whether men and women react similarly to FEEDBACK and VOTE, we estimate the main regressions separately for both subgroups with and without controls (see Table 5). Here we see that FEEDBACK has an economically and statistically significant effect on the quantity of work done by male volunteers, while the effect on females' motivation is smaller and less significant. The effect among men is sizeable: the prospect of getting feedback motivated the men in that treatment to color 4.2 cards, twice as many as the men in the control group colored. Even taking into account control variables, the coefficient tells us that the treatment results in an increase of 1.8 cards, which is equivalent to an increase of approximately 50 percent compared to the average man in the control group. Note that the promised feedback was provided anonymously, so explanations related to social image concerns (Ariely et al. 2009) should not play a role in here. Rather, men are known to be more susceptible to competitive environments than women (Gneezy and Rustichini 2000, Niederle and Vesterlund 2007). This finding is in line with previous findings from a lab-in-the-field experiment (Wozniak et al. 2016) showing that men seem more interested in comparing themselves to others, in that they are more likely than women to pay for costly feedback.¹⁸

¹⁸ We control for competitiveness in the gender estimations. The effects are persistent with and without the competitiveness variable. Therefore, we assume that the effect is not exclusively driven by competitiveness, but may be driven by a higher susceptibility to feedback on the part of the men.

Intriguingly, we find somewhat reversed effects regarding VOTE. The opportunity to participate in the decision-making process motivates women to color 1.256 more cards, or 1.5 when adding control variables. The performance increase, however, is only observable among women and not men (the coefficient is neither economically nor statistically significant). This effect is in line with findings which suggest that women particularly like their opinion to make a difference (Konrad et al. 2008). More research is needed to improve our understanding on gender differences of performance-enhancing effects based on self-determined work. These results suggest a potential pathway for nonprofit practitioners to manage their volunteers more effectively by providing gender-specific treatments.

4.3.2 Task enjoyment

The tasks at episodic events are often simple, repetitive, and monotonous. This constitutes an important prerequisite at episodic events where large numbers of heterogeneous supporters should be able to participate directly or after a very short training period. Besides, if there are different tasks with high and low levels of impact and autonomy,¹⁹ a recent literature review (Hyde et al. 2016) suggests that the tasks with higher degrees of impact, responsibility, and decision making are usually assigned to the regular volunteers who are familiar with the organization and who are expected to be more reliable. The drawback of simple, repetitive tasks is obvious: they are not likely to motivate volunteers in and of themselves. The reverse is likely true for interesting tasks, so we expected that the individual subjective perception of *task enjoyment* would be positively correlated with the quantity measure of performance, which proved to be the case ($p=0.022$). We split the participants into two groups using a median split on *task enjoyment* (1–7), with the median being 4.667.²⁰ The mean of colored cards among the low-enjoyment volunteers is 3.34, while the high-enjoyment individuals colored an average of 3.85 cards. Table 6 reveals that FEEDBACK and VOTE both have strong effects on low-enjoyment volunteers, while the high-enjoyment volunteers responded to VOTE but less to FEEDBACK. For the latter, the regression coefficient diminishes and is barely significant. Interpreting this result yields some relevant management implications. Every supervisor, whether of paid staff or volunteers, faces the issue of workers having low task enjoyment. The results suggest that empowerment may constitute a simple way to increase engagement for people with low enjoyment performing them.

¹⁹ Examples of basic tasks include preparing and providing food or information at events, putting fundraising letters in envelopes, and collecting money in door-to-door fundraising campaigns. More strategic work tasks include budgeting decisions, the evaluation of fundraising methods, decisions about how to spend money for needy persons, and planning of work assignments.

²⁰ In total, there were 16 volunteers with a median value of 4.667. They were added to the bottom 50% of task enjoyment. Adding them to the upper half does not change the results qualitatively.

Table 6 Heterogeneous treatment effects by task enjoyment

Specification	High task enjoyment (1)	Low task enjoyment (2)
Feedback	0.883* (0.460)	1.143*** (0.304)
Vote	1.372** (0.571)	1.278*** (0.327)
Controls	Yes	Yes
Constant	4.115 (3.086)	2.060* (1.104)
Observations	79	105
R ²	0.112	0.168

The table presents ordinary least squares estimates of treatment effects. In both specifications (1) and (2), the dependent variable is the total amount of colored cards. The reference group is CONTROL. The levels of *task enjoyment* were divided into two groups at the median. Controls included were age, gender, willingness to support, competitiveness. Robust standard errors are displayed in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.3.3 Competitiveness

Although competitiveness in voluntary tasks is not the primary focus of NPO managers, its massive influence on workplace behavior has been impressively shown (Niederle and Vesterlund 2007; Buser et al. 2014; Banuri et al. 2018). Differences in competitiveness can explain differences in education, labor market outcomes, and scientific careers (Buser et al. 2014), and it is also strongly correlated with the effectiveness of relative performance feedback (Banuri et al. 2018). We estimate the standard regression with total colored cards as the dependent variable once with the upper median bound of competitiveness, and once with the lower median bound (see Table 7). While FEEDBACK has a strong positive effect on total colored cards for volunteers with above-median competitiveness (coefficient = 1.162, $p = 0.002$), it does motivate volunteers with low levels of competitiveness less (coefficient = 0.031). Hence, the results suggest that granting costless individual performance feedback is useful in the aggregate among volunteers, and especially among those with high levels of competitiveness.

Table 7 Heterogeneous treatment effects by competitiveness

Specification	High competitiveness (1)	Low competitiveness (2)
Feedback	1.162*** (0.357)	0.946* (0.476)
Vote	0.890** (0.407)	1.582*** (0.428)
Controls	Yes	Yes
Constant	0.467 (1.482)	3.280 (2.226)
Observations	98	84
R ²	0.161	0.179

The table presents ordinary least squares estimates of treatment effects. In both specifications (1) and (2), the dependent variable is the total amount of colored cards. The reference group is CONTROL. The levels of *competitiveness* were divided at the median into two groups. Controls included were age, gender, willingness to support, task enjoyment. Robust standard errors are displayed in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

5 Conclusion

In this paper, we study the performance, and the potential boosters thereof, of unpaid volunteers using a collaborative field experiment. This is novel from at least three perspectives. First, we focus on actual work behavior. This is an important outcome among paid workers and has been researched accordingly. But despite its importance for the functioning of the nonprofit sector, focusing on the actual work performance of volunteers is surprisingly rare in research and practice alike. Second, we derive inspiration from self-determination theory on how to support volunteers' intrinsic motivation, given that volunteering is too different from paid labor to apply classic monetary or non-monetary incentive schemes: those schemes either tend to fail in such a context or are simply inappropriate for volunteers. The treatments we implemented are novel in research and easy to put into practice even in the presence of budget constraints. Third, we can reliably investigate causal effects by using a randomized and controlled environment, a rather unusual approach when studying volunteer motivation. The randomized control is not compromised by the costs of artificial environment (Harrison and List 2004; List 2011) as our experiment was conducted in cooperation with a charity under natural circumstances.

Both the FEEDBACK and VOTE treatments boost performance. We observe a large increase in the average total number of colored cards per volunteer through the announcement of individual performance feedback (36.5%) and an even larger increase (44.6%) through the opportunity to vote on the donation use. As their implementation does not involve significant costs, but raises average performance without any decrease in quality, both instruments are effective for charities in managing their human resources more efficiently. The rich data set allowed us to investigate treatment effects on volunteers' individual performance and subgroup effects.

Our study is only a first attempt at using a natural field experiment to investigate heterogeneous treatment effects in the domain of volunteer motivation. If the preliminary results are confirmed by further studies, nonprofit organizations can motivate their volunteers more efficiently by providing specific treatments to different target groups. Our results suggest that the men significantly increase their performance by the opportunity to receive individual performance feedback, while the women increase their effort when provided with the opportunity to vote. Importantly, the treatments did not affect the quality of the work, even though the volunteers were encouraged to maximize both quantity and quality. The information that customers would pay whatever they wished for the greeting cards made it clear that it was very important to produce beautiful cards. In fact, every worker seems to have had a specific quality level that could not be exogenously manipulated, at least not by the instruments that we examined, and obviously not in the short run.

While episodic volunteers were the focus of this paper, it would be exciting to study these instruments among highly intrinsically motivated paid workers—for example in geriatric care, in youth health, in nursing, or in services for homeless people, addicted people, or pregnant women. Further, the tools proved to be particularly effective for a workforce with low task enjoyment, so our findings could also be relevant for paid one-time or casual jobs where commitment to the organization is low and people tend to be less intrinsically motivated—for example in classic day labor (e.g. in developing countries and/or the manufacturing industry), in digital day labor (Uber, Mechanical Turk, Foodora), or in agricultural labor.

To the best of our knowledge, this is the first paper providing field experimental evidence on work motivation in an episodic volunteering setting. We conclude that more research is needed to understand work motivation in volunteerism, as “one swallow doesn’t make a summer” (Maniadis et al. 2014). We left out some related topics that deserve similar attention, such as how particular arrangements and work characteristics of voluntary settings affect volunteers’ willingness to engage in an episodic event, or the probability that people will volunteer regularly for a social organization to maintain the social capital of our society.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Since 2001, Kinderherzhilfe has been engaged in supporting children who are born with heart disease. The acknowledged charity supports not only the affected children but also their parents and siblings through information and direct support, as well as solely through distraction of the problems that the illness brings. **The charity is therefore in need of donations to finance activities like barbecues, holiday trips, information brochures, and support for children’s wishes and toys for the cardiac wards in hospitals.** For the project in support of Kinderherzhilfe, we have designed a short questionnaire. We appreciate your taking the time to fill it in. All information is of course anonymous.

Note: The variable *willingness to support* includes all 6 items from Handout 1 (Cronbach’s alpha 0.71).

<i>We are interested in your opinion towards Kinderherzhilfe. How do the following statements apply to you personally?</i>	Strongly disagree							Strongly agree
The value 1 means: strongly disagree The value 7 means: strongly agree	1	2	3	4	5	6	7	
I’m familiar with the Kinderherzhilfe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supporting children with heart disease is important to me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I could imagine contributing financially.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I could imagine volunteering for the Kinderherzhilfe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I could imagine informing my friends about Kinderherzhilfe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I could imagine promoting Kinderherzhilfe to strangers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 7 Handout I (english version)

CONTROL	FEEDBACK	VOTE
<p>All the raised during this event will be for the Kinderherzhilfe. To make this event a success, we need your help!</p> <p>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can. Both quantity and quality are important because the produced cards will be sold using a pay-what-you-want mechanism. That means that customers will select the greeting cards they like the most and pay the price they like. Therefore, we need many and beautifully colored greeting cards because we are hoping to raise a high amount of money this way. At the end of the activity, the money raised will be given to Kinderherzhilfe.</p> <p>We want to support Kinderherzhilfe together. In light of this, we thank you for your involvement.</p>	<p>All the raised during this event will be for the Kinderherzhilfe. To make this event a success, we need your help!</p> <p>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can. Both quantity and quality are important because the produced cards will be sold using a pay-what-you-want mechanism. That means that customers will select the greeting cards they like the most and pay the price they like. Therefore, we need many and beautifully colored greeting cards because we are hoping to raise a high amount of money this way. At the end of the activity, the money raised will be given to Kinderherzhilfe.</p> <p>We want to support Kinderherzhilfe. together. In light of this, we thank you for your involvement.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>After the event, there will be an anonymous list saying how much each one of you earned for the good cause. For this, you will need your supporter number. Please keep it.</p> </div>	<p>All the raised during this event will be for the Kinderherzhilfe. To make this event a success, we need your help!</p> <p>We ask you to color as many and as beautiful mandalas with the provided pastel crayons as you can. Both quantity and quality are important because the produced cards will be sold using a pay-what-you-want mechanism. That means that customers will select the greeting cards they like the most and pay the price they like. Therefore, we need many and beautifully colored greeting cards because we are hoping to raise a high amount of money this way. At the end of the activity, the money raised will be given to Kinderherzhilfe.</p> <p>We want to support Kinderherzhilfe together. In light of this, we thank you for your involvement.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>As a supporter, you are allowed to vote for which charitable purpose the collected money will be spent. Please use the ballot at the bottom to vote.</p> </div>

Ballot (*only Vote treatment*)

<p>The money raised will be used for the purpose that you and the other participants find best. For which exact purpose should the raised money be used? You are allowed to vote. You may select multiple activities as desired.</p>	Barbecue	Holiday trip	Printing flyers	Individual wishes of children	Toys for the cardiac ward at hospitals
---	----------	--------------	-----------------	-------------------------------	--

Fig. 8 Handout II—treatments (english version)

Gender: Female: <input type="checkbox"/> Male: <input type="checkbox"/> Age: _____ Support ID: _____
Dear participant, We are a group of students who have designed a questionnaire for a project supporting Kinderherzhilfe. We thank you for taking the time to fill it out.

<i>To what extent do the following statements apply to you personally?</i>	<i>Does not apply at all</i>							<i>Fully applies</i>
<i>The value 1 means: does not apply at all</i> <i>The value 7 means: fully applies</i>	1	2	3	4	5	6	7	
1. I enjoyed coloring mandalas.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2. I would like to color mandalas again on occasion.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3. I feel better after coloring than before.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4. I always have the ambition to be better than average.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5. I am only satisfied when my performance is above average.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6. To be successful in life means to me to be better than others.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
7. I would like to be one of the best in all areas of life (job, school, sports, etc.).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Note: Statements 1-3 were related to volunteers' task enjoyment (Cronbach's alpha = 0.88) while statements 4-7 were related to volunteers' competitiveness (Cronbach's alpha = 0.77).

Fig. 9 Questionnaire

Table 8 Main regressions (with quality measures)

Specification	Total painted cards				
	(1)	(2)	(3)	(4)	(5)
Feedback	1.027*** (0.262)	0.969*** (0.263)	1.125*** (0.262)	0.807*** (0.229)	0.610*** (0.193)
Responsibility	1.254*** (0.300)	1.182*** (0.297)	1.300*** (0.302)	1.364*** (0.272)	1.312*** (0.225)
Age		-0.0117 (0.0487)	-0.0105 (0.0509)	0.0206 (0.0461)	0.0487 (0.0404)
Gender		0.478* (0.281)	0.430 (0.310)	0.759*** (0.272)	0.677*** (0.213)
Willingness to support			0.0373 (0.129)	0.0493 (0.122)	-0.0434 (0.0970)
Task enjoyment			0.177** (0.0748)	0.226*** (0.0658)	0.149*** (0.0516)
Competitiveness			-0.0661 (0.0862)	-0.0504 (0.0798)	-0.0411 (0.0664)
Rating				-0.821*** (0.147)	-1.001*** (0.130)
Colors				-0.372*** (0.108)	-0.129 (0.0817)
Areas				-0.337 (0.275)	-0.287 (0.197)
Sold cards					0.669*** (0.108)
Sales rank					-0.000675 (0.000831)
Sales					-0.00598 (0.0710)
Constant	2.810*** (0.174)	2.771** (1.150)	2.039 (1.254)	6.496*** (1.697)	4.738*** (1.374)
Observations	187	185	181	180	180
R ²	0.102	0.112	0.157	0.352	0.547

The table presents ordinary least square estimates of treatment effects. In all five reported columns from (1)–(5), the dependent variable is the total amount of colored cards. The reference group in all estimations is CONTROL. *Willingness to support* is the mean of six questions concerning someone's willingness to support the charity. *Task enjoyment* is also a mean of three task-related enjoyment statements of drawing mandalas. *Competitiveness* is the mean of four related statements *Rating* is the mean assessment of volunteers' painted cards. *Colors* is the amount of volunteer' mean colors used per mandala. *Areas* is the mean amount of areas someone colored per mandala. *Sold cards* is the number of total sold cards. *Sales rank* is the ordering in which the cards got sold. *Sales* are the total earnings volunteers earned for their painted cards. Robust standard errors are displayed in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9 Additional descriptive statistics

Variable	Observations	Mean	SD	Min	Max
Cards	187	3.567	1.716	0	10
Rating	185	3.214	0.646	1	4.67
Colors	185	4.708	1.040	1.8	8.5
Areas	185	4.639	4.639	1.5	5.5
Sold cards	185	1.745	1.262	0	7
Sales rank	185	115.089	94.018	0	333
Sales	185	1.997	1.745	0	10

All values are calculated at the individual level

Table 10 Quality regressions with and without controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Rating	Rating	Colors	Colors	Areas	Areas	Sold cards	Sold cards	Sales rank	Sales rank	Sales	Sales
FEEDBACK	-0.249** (0.109)	-0.198* (0.109)	-0.351* (0.196)	-0.308 (0.205)	-0.0939 (0.101)	-0.0902 (0.110)	0.305 (0.215)	0.344 (0.221)	-20.68 (17.52)	-17.79 (18.17)	0.288 (0.308)	0.300 (0.326)
VOTE	0.103 (0.115)	0.0838 (0.116)	-0.0297 (0.180)	-0.0348 (0.193)	-0.0867 (0.0828)	-0.112 (0.0905)	0.0484 (0.231)	0.104 (0.237)	-19.37 (16.67)	-18.14 (17.51)	-0.0795 (0.316)	0.0347 (0.345)
Age		0.00801 (0.0134)	0.0492** (0.0214)			0.00373 (0.0147)		-0.0603** (0.0258)		-1.560 (2.490)		-0.0960** (0.0438)
Gender		0.317*** (0.108)	-0.0380 (0.207)		0.186 (0.115)		0.202 (0.229)		-6.272 (16.70)		0.0817 (0.038)	
Willingness to support		-0.0254 (0.0416)	0.0508 (0.0818)		-0.0117 (0.0398)		0.112 (0.102)		-2.734 (7.058)		0.0788 (0.155)	
Task enjoyment		0.0655** (0.0254)	0.0418 (0.0502)		-0.000862 (0.0240)		0.120** (0.0544)		2.341 (3.876)		0.118 (0.0774)	
Competitiveness		0.00662 (0.0292)	0.0418 (0.0492)		-0.00989 (0.0291)		-0.0266 (0.0638)		0.119 (4.928)		-0.0166 (0.104)	
Constant	2.830*** (0.0748)	2.198*** (0.392)	4.834*** (0.137)	3.235*** (0.607)	4.699*** (0.0637)	4.572*** (0.414)	1.629*** (0.149)	1.129* (0.586)	130.9*** (12.48)	163.8*** (66.58)	1.931*** (0.215)	3.198*** (0.584)
Observations	185	180	185	180	185	180	185	180	187	181	187	181
R ²	0.052	0.118	0.023	0.051	0.007	0.032	0.011	0.328	0.012	0.012	0.008	0.320

The table presents ordinary least square estimates for twelve regressions (six with and six without control variables) for the dependent variables *rating*, *colors*, *areas*, *sold cards*, *sales rank*, and *sales*

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 11 Multiple hypothesis testing

Outcome	Comparison	Difference in means	<i>p</i> -values			
			Unadj		Multipl. adj	
			Remark 3.1	Thm 3.1	Bonferroni	Holm
Total cards	Control vs. Feedback	1.0265	.0003	.0003	.0047	.0043
Total cards	Control vs. Vote	1.2539	.0003	.0003	.0047	.0047
Rating	Control vs. Feedback	.24864	.0240	.1993	.3360	.2880
Rating	Control vs. Vote	.10284	.3646	.7363	1.000	1.000
Colors	Control vs. Feedback	.35061	.0760	.4773	1.000	.836
Colors	Control vs. Vote	.02966	.8757	.9867	1.000	1.000
Areas	Control vs. Feedback	.09390	.3583	.8187	1.000	1.000
Areas	Control vs. Vote	.08669	.2840	.7907	1.000	1.000
Sold cards	Control vs. Feedback	.30539	.1673	.7423	1.000	1.000
Sold cards	Control vs. Vote	.04838	.8330	.9743	1.000	1.000
Sales rank	Control vs. Feedback	2.0676	.2473	.8377	1.000	1.000
Sales rank	Control vs. Vote	1.9370	.2630	.8260	1.000	1.000
Sales	Control vs. Feedback	.28926	.1890	.7860	1.000	1.000
Sales	Control vs. Vote	.03225	.8887	.8887	1.000	.8887

The table presents *p*-values for two-sided comparisons of means according to the multiple hypothesis tests procedure of List et al. (2019)

References

- Al-Ubaydli, O., & Lee, M. (2011). Can tailored communications motivate environmental volunteers? A natural field experiment. *American Economic Review*, *101*(3), 323–328.
- Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy*, *97*(6), 1447–1458.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, *100*(401), 464–477.
- Andreoni, J. (2006). Philanthropy. *Handbook of the Economics of Giving, Altruism and Reciprocity*, *2*, 1201–1269.
- Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization*, *67*(3), 671–677.
- Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, *99*(1), 544–555.
- Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, *120*, 1–17.
- Australian Bureau of Statistics (2015). *General Social Survey: Summary Results*, Australia, 2014.
- Azmat, G., & Iriberry, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, *94*(7), 435–452.
- Azmat, G., Bagues, M., Cabrales, A., & Iriberry, N. (2019). What you don't know... can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science* (forthcoming).
- Bäker, A., & Mechtel, M. (2018). The role of task meaning on output in groups: Experimental evidence. *Managerial and Decision Economics*, *39*(2), 131–141.
- Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, *34*, 13–25.

- Banuri, S., Dankova, K., & Keefer, P. (2018). *It's not all fun and games: Feedback, task motivation, and effort*. Working Paper No. 17–10. School of Economics, University of East Anglia, Norwich, UK.
- Barbara, L., Grolleau, G., & Mzoughi, N. (2017). Do you prefer having more or more than others in the workplace? A quasi-experimental survey in Algeria. *Managerial and Decision Economics*, 38(4), 595–606.
- Bauer, T. K., Bredtmann, J., & Schmidt, C. M. (2013). Time versus money—The supply of voluntary labor and charitable donations across Europe. *European Journal of Political Economy*, 32, 80–94.
- Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5), 1652–1678.
- Beder, J., & Fast, J. (2008). Episodic volunteering: Why people walk/run for charity. *The International Journal of Volunteer Administration*, 25(3), 3–13.
- Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, 40(5), 924–973.
- Bidee, J., Vantilborgh, T., Pepermans, R., Huybrechts, G., Willems, J., Jegers, M., & Hofmans, J. (2013). Autonomous motivation stimulates volunteers' work effort: A self-determination theory approach to volunteerism. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 24(1), 32–47.
- Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, 50(2), 368–425.
- Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11), 3085–3099.
- Brown, A. L., Meer, J., & Williams, J. F. (2019). Why do people volunteer? An experimental analysis of preferences for time donations. *Management Science*, 65(4), 1455–1468.
- Bureau of Labor Statistics (2016). *Volunteering in the United States—2015*. Washington, D.C.: U.S. Department of Labor.
- Burns, D. J., Reid, J. S., Toncar, M., Fawcett, J., & Anderson, C. (2006). Motivations to volunteer: The role of altruism. *International Review on Public and Nonprofit Marketing*, 3(2), 79–91.
- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3), 1409–1447.
- Cardinaels, E., Chen, C. X., & Yin, H. (2018). Leveling the playing field: The selection and motivation effects of tournament prize spread information. *The Accounting Review*, 93(4), 127–149.
- Carpenter, J., & Myers, C. K. (2010). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics*, 94(11), 911–920.
- Carpenter, J., & Gong, E. (2016). Motivating Agents: How much does the mission matter? *Journal of Labor Economics*, 34(1), 211–236.
- Chadi, A., Jeworrek, S., & Mertins, V. (2017). When the meaning of work has disappeared: Experimental evidence on employees' performance and emotions. *Management Science*, 1696–1707.
- Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, 90, 123–133.
- Charness, G., Cobo-Reyes, R., Jimenez, N., Lacombe, J. A., & Lagos, F. (2012). The hidden advantage of delegation: Pareto improvements in a gift exchange game. *American Economic Review*, 102, 2358–2379.
- Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., et al. (1998). Understanding and assessing the motivations of volunteers: A functional approach. *Journal of Personality and Social Psychology*, 74(6), 1516.
- Cnaan, R. A., & Handy, F. (2005). Towards understanding episodic volunteering. *Vrijwillige Inzet Onderzoek*, 2(1), 29–35.
- Conrads, J., Irlenbusch, B., Reggiani, T., Rilke, R. M., & Sliwka, D. (2016). How to hire helpers? Evidence from a field experiment. *Experimental Economics*, 19(3), 577–594.
- Crosan, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–474.
- Culp, K., & Nolan, M. (2000). Trends impacting volunteer administrators in the next ten years. *Journal of Volunteer Administration*, 19(1), 10–19.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109–134.
- Dunn, J., Chambers, S. K., & Hyde, M. K. (2016). Systematic review of motives for episodic volunteering. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(1), 425–464.

- Eckstein, S. (2001). Community as gift-giving: Collectivistic roots of volunteerism. *American Sociological Review*, 66, 829–851.
- Ellingsen, T., & Johannesson, M. (2009). Time is not money. *Journal of Economic Behavior & Organization*, 72(1), 96–102.
- Englert, B., & Helmig, B. (2018). Volunteer performance in the light of organizational success: A systematic literature review. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 29(1), 1–28.
- Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6), 679–688.
- Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1), 39–57.
- Franke, J., Gurtoviy, R., & Mertins, V. (2016). Workers' participation in wage setting: A gift-exchange experiment. *Journal of Economic Psychology*, 2016, 151–162.
- Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion*, 27(3), 199–223.
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331–362.
- Gerhards, L., & Siemer, N. (2015). The impact of private and public feedback on worker performance: Evidence from the lab. *Economic Inquiry*, 54(2), 1188–1201.
- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3), 791–810.
- Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637–1664.
- Goette, L., Stutzer, A., & Frey, B. M. (2010). Prosocial motivation and blood donations: a survey of the empirical literature. *Transfusion Medicine and Hemotherapy*, 37(3), 149–154.
- Grouzet, F. M., Vallerand, R. J., Thill, E. E., & Provencher, P. J. (2004). From environmental factors to outcomes: A test of an integrated motivational sequence. *Motivation and Emotion*, 28(4), 331–346.
- Haivas, S. M., Hofmans, J., & Pepermans, R. (2012). Self-determination theory as a framework for exploring the impact of the organizational context on volunteer motivation: A study of Romanian volunteers. *Nonprofit and Voluntary Sector Quarterly*, 41(6), 1195–1214.
- Handy, F., Brodeur, N., & Cnaan, R. A. (2006). Summer on the island: Episodic volunteering. *Voluntary Action*, 7(3), 31–46.
- Harrison, D. A. (1995). Volunteer motivation and attendance decisions: Competitive theory testing in multiple samples from a homeless shelter. *Journal of Applied Psychology*, 80(3), 371.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009–1055.
- Heinz, M., & Schumacher, H. (2017). Signaling cooperation. *European Economic Review*, 98, 199–216.
- Harris, P. B., & Houston, J. M. (2010). A reliability analysis of the revised competitiveness index. *Psychological Reports*, 106(3), 870–874.
- Hustinx, L. (2001). Individualisation and new styles of youth volunteering: An empirical exploration. *Voluntary Action*, 3(2), 57–76.
- Hustinx, L., & Lammertyn, F. (2003). Collective and reflexive styles of volunteering: A sociological modernization perspective. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 14(2), 167–187.
- Hustinx, L., Haski-Leventhal, D., & Handy, F. (2008). One of a kind? Comparing episodic and regular volunteers at the Philadelphia Ronald McDonald House. *International Journal of Volunteer Administration*, 25(3), 50–66.
- Hyde, M. K., Dunn, J., Scuffham, P. A., & Chambers, S. K. (2014). A systematic review of episodic volunteering in public health and other contexts. *BMC Public Health*, 14, 992.
- Hyde, M. K., Dunn, J., Bax, C., & Chambers, S. K. (2016). Episodic volunteering and retention: An integrated theoretical approach. *Nonprofit and Voluntary Sector Quarterly*, 45(1), 45–63.
- Irlenbusch, B., & Ruchala, G. K. (2008). Relative rewards within team-based compensation. *Labour Economics*, 15(2), 141–167.
- Jeworrek, S., & Mertins, V. (2019a). Mission, motivation, and the active decision to work for a social cause. Halle Institute for Economic Research. Discussion Papers (No. 10).
- Jeworrek, S., & Mertins, V. (2019b). Wage delegation in the field. *Journal of Economics & Management Strategy*. <https://doi.org/10.1111/jems.12313>.
- Konrad, A. M., Kramer, V., & Erkut, S. (2008). The impact of three or more women on corporate boards. *Organizational Dynamics*, 37(2), 145–164.

- Kosfeld, M., & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3), 86–99.
- Kosfeld, M., Neckermann, S., & Yang, X. (2017). The effects of financial and recognition incentives across work contexts: The role of meaning. *Economic Inquiry*, 55(1), 237–247.
- Levitt, S., & Neckermann, S. (2014). What field experiments have and have not taught us about managing workers. *Oxford Review of Economic Policy*, 30(4), 639–657.
- List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, 25(3), 3–16.
- List, J. A., & Rasul, I. (2010). Field experiments in labor economics. In D. Card & O. Ashenfelter (Eds.), *Handbook of labor economics (Vol. 4a)*. Amsterdam, North Holland.
- List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4), 773–793.
- Maniadi, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1), 277–290.
- Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: was Titmuss right? *Journal of the European Economic Association*, 6(4), 845–863.
- Meer, J. (2014). Effects of the price of charitable giving: Evidence from an online crowdfunding platform. *Journal of Economic Behavior & Organization*, 103, 113–124.
- Merrill, M. (2006). Global trends and the challenges for volunteering. *The International Journal of Volunteer Administration*, XXIV(1), 9–14.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 133(3), 1067–1101.
- Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *The Journal of Economic Perspectives*, 24(2), 129–144.
- Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–78.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Smith, K., Holmes, K., Haski-Leventhal, D., Cnaan, R. A., Handy, F., & Brudney, J. L. (2010). Motivations and benefits of student volunteering: Comparing regular, occasional, and non-volunteers in five countries. *Canadian Journal of Nonprofit and Social Economy Research*, 1(1), 65–81.
- Smither, R. D., & Houston, J. M. (1992). The nature of competitiveness: The development and validation of the competitiveness index. *Educational and Psychological Measurement*, 52(2), 407–418.
- Tonin, M., & Vlassopoulos, M. (2010). Disentangling the sources of pro-socially motivated effort: A field experiment. *Journal of Public Economics*, 94(11), 1086–1092.
- Turcotte, M. (2015). *Volunteering and charitable giving in Canada* (pp. 89–652). Spotlight on Canadians: Result from the General Social Survey. Statistics Canada. Catalogue no. 176–212.
- Wozniak, D., Harbaugh, W., & Mayr, U. (2016). *The effect of feedback on gender differences in competitive choices* (p. 1976073). ID: SSRN Scholarly Paper No.
- Ziemek, S. (2006). Economic analysis of volunteers' motivations—A cross-country study. *The Journal of Socio-Economics*, 35(3), 532–555.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.