Universität
Basel

Fakultät für
Psychologie

Social Preferences Across Contexts

Inauguraldissertation zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der Fakultät für Psychologie der Universität Basel von

Dominik Meier

aus Emmen

Basel, 2023

Genehmigt von der Fakultät für Psychologie auf Antrag von

Prof. Dr. Jörg Rieskamp

Prof. Dr. Georg von Schnurbein

Datum des Doktoratsexamen: 30.08.2023

Dekan:in der Fakultät für Psychologie

Universität Basel

Fakultät für Psychologie

Erklärung zur wissenschaftlichen Lauterkeit

Ich erkläre hiermit, dass die vorliegende Arbeit ohne die Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel selbstständig verfasst habe. Zu Hilfe genommene Quellen sind als solche gekennzeichnet. Die veröffentlichten oder zur Veröffentlichung in Zeitschriften eingereichten Manuskripte wurden in Zusammenarbeit mit den Koautoren erstellt und von keinem der Beteiligten an anderer Stelle publiziert, zur Publikation eingereicht, oder einer anderen Prüfungsbehörde als Qualifikationsarbeit vorgelegt. Es handelt sich dabei um folgende Manuskripte:

1. Meier, D.S., Rieskamp, J., & Schöbel, M. (2022). Why do People Avoid Sharing Situations? Testing Guilt Aversion Versus Self-Image Concerns.

2. Meier, D.S. (2023). Compassion for all: Real-World Online Donations Contradict Compassion Fade.

3. Meier, D.S., Petrig, A., & von Schnurbein, G. (in press). Risking Your Health to Help Others: Volunteering During the COVID-19 Pandemic. *Nonprofit and Voluntary Sector Quarterly.*

Spezifizierung des eigenen Forschungsbeitrags zu den Manuskripten:

1. Eigener Beitrag nach CRediT[1]:

[1] https://casrai.org/credit/

☒ Conceptualization ☒ Data curation Formal ☒ Analysis
☐ Funding acquisition ☒ Investigation ☒ Methodology
☒ Project administration ☒ Resources Software ☒
☐ Supervision ☒ Validation ☒ Visualization
☒ Writing – original draft
☒ Writing – review & editing

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

2. Eigener Beitrag nach CRediT[1]:

☒ Conceptualization ☒ Data curation Formal ☒ Analysis
☐ Funding acquisition ☒ Investigation ☒ Methodology
☒ Project administration ☒ Resources Software ☒
☐ Supervision ☒ Validation ☒ Visualization
☒ Writing – original draft
☒ Writing – review & editing

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

3. Eigener Beitrag nach CRediT[1]:

☒ Conceptualization ☒ Data curation Formal ☒ Analysis
☐ Funding acquisition ☒ Investigation ☒ Methodology
☒ Project administration ☒ Resources Software ☒
☐ Supervision ☒ Validation ☒ Visualization
☒ Writing – original draft
☒ Writing – review & editing

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

Open-Science Aspekte der Manuskripte:

Universität Basel

Fakultät für Psychologie

1. Preregistration: nein ☐ ja ☒

   Open-Access-Publikation: ☐ ja ☒ nein

   Open-Access- ☐ ja ☒ Data/Analyse: nein

   Ort/URL der Daten und Analysen: …………………………………..

2. Preregistration: nein ☐ ja ☒

   Open-Access-Publikation: ☐ ja ☒ nein

   Open-Access- ☐ ja ☒ Data/Analyse: nein

   Ort/URL der Daten und Analysen: …………………………………..

3. Preregistration: nein ☐ ja ☒

   Open-Access-Publikation: ☐ ja ☒ nein

   Open-Access- ☐ ja ☒ Data/Analyse: nein

   Ort/URL der Daten und Analysen: …………………………………..

Ort, Datum ………………………………………………….

Signatur ………………………………………………….

Vorname Nachname ………………………………………………….

Acknowledgment

Although doing a PhD can be arduous, I thoroughly enjoyed this journey most of the time. I deeply value doing something I genuinely enjoy, not least because my parents did not always share this privilege. Therefore, I am immensely grateful to them for providing me with opportunities they did not have access to and for all the support they provided along the way. My deepest gratitude also extends to Lara. I could not imagine a better partner not just for this journey, but for the journey that is life. Although you often said that you didn't know what I was doing (for which I take the blame), you always sensed how I was doing and offered the best support anyone could hope for. I'm also grateful to my friends and colleagues who supported me on this journey. Thanks, Gregor, for our great discussions during long gravel rides and coffee breaks in your office. Thanks, Dimitris, for being such a great office mate. Your temperament always cheered me up and I always enjoyed our sometimes more and sometimes less serious discussions. I also want to thank Adrian for taking me on his triathlon journey. Having you as a training partner greatly increased my work-life balance at the beginning of my PhD and made even the longest run or bike ride enjoyable.

I am also thankful to my current and former colleagues, many of whom became friends over the last four years. Thanks Lucca, Noam, Alice, Anja, Mira and Carina for the great company and the regular walks around the block that often lead to fruitful ideas. The same gratefulness also extends to Sara, Sophie, Theresa, Oto, Damian, Nicholas and other former or current members of the CEPS. You made the first part of this journey as enjoyable and memorable as the second. Once a week I had the pleasure of visiting my office at the psychology faculty, and it was because of people like Kevin, Florian, Loris, Nicolas, Zak et al. that I always looked forward to that day. Special thanks also go to Rui and Dirk. Collaborating with you on non-PhD related topics was as much fun as it was interesting and insightful.

Finally, I want to thank those who most closely support my PhD. First and foremost, I want to thank my advisors, Jörg Rieskamp and Georg von Schnurbein, for their invaluable support on this journey. They gave me great intellectual freedom but also provided, in the form of great feedback and advice, the guards and rails to make sure I don’t get lost along the way. I also want to thank Steve for the great advice he provided along the way. Especially the careerrelated discussions at the beginning of my PhD were very enlightening for me as a “first-gen”.

Last but not least, I want to thank Markus. I could not have hoped for a better mentor.

Abstract

Over the past decades, the behavior of people who do not maximize their payoff but instead seem to be other-regarding has received much attention in the (behavioral) economics literature. Many different social preference models that ideally explain such other-regarding behavior across a large span of contexts have been proposed and tested. Building on this literature, this dissertation studies social preferences in different contexts and expressions across three manuscripts. The first manuscript examines the behavior of people who avoid a situation that allows them to express social preferences. Drawing on psychological game theory, we tested whether guilt-aversion or self-image concerns could better explain this behavior. It was found that guilt-aversion, but not self-image concerns, can explain the behavior of these people. The second manuscript made use of crowdfunding donations data and showed that the reversal of the compassion fade effect when going from a separate to a joint evaluation condition extends from the lab to the field. Social preferences can also manifest themselves through people donating their time. The third manuscript examines how the severity of a catastrophe (i.e., the COVID-19 pandemic) affects the provision of catastrophe-related voluntary labor. We found a concave relationship between the weekly COVID-19-related death numbers and the amount of voluntary work provided by individuals. Thus, by drawing on prosocial behavior expressed in three different environments, this dissertation extends the current literature by studying how social preferences are influenced by the context in which they are carried out.

How Social Preferences are Influenced by and Change Across Contexts

> Charity collector: 'I want you to give me a pound, and then I go away and give it to the orphans.'

Merchant banker: ‘Yes.’

Charity collector: ‘Well, that’s it.’

Merchant banker: ‘No, no, no! I don’t follow this at all. I mean, I don’t want to seem stupid, but it looks to me as though I’m a pound down on the whole deal.’

Monty Python’s merchant banker sketch

One does not need to be a merchant banker to see that from a purely monetary perspective, donating money to a charitable cause seems like a bad deal. But why then do so many people give to charity? This question has puzzled economists for the past decades, as the assumption that people act as if selfish is at the core of the standard economic model. The quest to study why people do not seem to follow this assumption too closely started with the study of Güth et al. (1982), who reported “the kind of empirical finding that surprises only economists” (Camerer, 2011, p. 53). Güth et al. (1982) introduced the “ultimatum game” where a “proposer” divides some amount of money between herself and another person. The other person can accept or reject the proposed split, in which case both will leave the game empty-handed. To most economists’ surprise, they found that unfair proposals were often rejected. This has been extensively replicated in studies that followed Güth et al. (1982), which found that on average 40 percent of the money is offered by the proposers, and small offers of 20 percent are rejected half the time (Camerer, 2011). These findings posed a puzzle to economists of this time, as almost all economic models from that time assumed that “*all* people are *exclusively* pursuing their material self-interest and do not care about ‘‘social’’ goals per se” (Fehr & Schmidt, 1999, p. 817).

But maybe proposers in the Ultimatum game make generous offers not because they care about the monetary payoff of the recipient (altruism) but because they anticipate that low offers will be rejected. Given this fear of rejection of low offers, it is rational for proposers to make (generous) offers that will not be rejected. To test whether generous proposers act out of fear or altruism, recipients have been stripped of their right to reject low offers in the "dictator game". Kahneman, Knetsch, and Thaler (1986) gave subjects the choice of splitting $20 with another subject either evenly ($10, $10) or unevenly ($18, $2). Three-quarters chose the equal split ($10, $10). Forsythe et al. (1994) directly compared offers in an ultimatum game with offers in a dictator game. If participants who make nonzero offers in these games are solely motivated by fairness concerns, offers should be the same in both games. However, offers in the dictator game were significantly lower, suggesting that offers in the Ultimatum game are partly motivated by strategic concerns (i.e., fear of rejection). But mean offers in the dictator game treatments of Forsythe et al. (1994) were still about 20% of the endowment, which is in line with pure altruism (Camerer, 2003, p. 66).

The early results from dictator games have been extensively replicated and extended. It is now undisputed that humans are systematically more benevolent than assumed by the standard economic model (i.e., *homo oeconomicus*) (Engel, 2011). In a meta-analysis conducted by Engel (2011) based on 616 treatments from other studies, dictators gave the recipient on average 28.35% of their endowment. But all is not lost for homo oeconomicus, as the most popular allocation decision (36.1%) was to give nothing to the recipient. However, a theory that can only explain about one-third of the participants' behavior is clearly not satisfactory. It therefore did not take long for economists to develop theories that can explain the prosocial behavior observed in the Dictator and the Ultimatum game. The prosocial behavior of the subjects in these games can be explained by assuming that people not only get utility from their own payoff but also from the payoff of others (Fehr & Schmidt, 1999). Although they all follow this same basic idea, a range of social preference models exist (Daruvala, 2010), and I will shortly summarize the most influential ones.

An early form of such a utility function was the pure altruism model, which assumes that an individual ! derives utility from private consumption "! and from e.g., the charity's output #, $("!, #) (Becker, 1974; Ottoni-Wilhelm et al., 2017). Since according to this model, individual ! gets utility from the total output #, donations by herself and others are perfect substitutes. Increases in giving by others should therefore crowd-out (i.e., decrease) individual !'s contribution dollar for dollar (Ottoni-Wilhelm et al., 2017). However, in contrast to this prediction, empirically observed crowding-out is relatively small (Abrams & Schitz, 1978; De Wit & Bekkers, 2017). This shortfall of the pure altruistic model was met with the concept of impure altruism (Andreoni, 1990), which assumes that a donor not only cares about the total contributions (#), but also about her own contribution ((!). The utility of one's own contribution is often called "warm-glow" (Andreoni, 1990). Andreoni (1990) showed that a model where utility is modeled in this way (i.e.,)!("!, #, (!)) can lead to incomplete crowding out. Future research has corroborated the importance of this additional warm-glow parameter (Ottoni-Wilhelm et al., 2017).

The "inequality-aversion" model proposed by Fehr and Schmidt (1999) is another highly influential account. Their model assumes that next to their payoffs, people also care about the difference between their payoff and the payoff of others. Their model has parameters for how much players dislike having less than others (envy) and dislike having more than others (guilt) (Camerer, 2011). At around the same time, Bolton and Ockenfels (2000) proposed a similar model of inequality aversion. In contrast to the model by Fehr and Schmidt (1999), the equity, reciprocity, and competition (ERC) model proposed by Bolton and Ockenfels (2000) assumes that people care about relative allocations, whereas the Fehr and Schmidt (1999) model assumes that people care about absolute differences. We could go on about the granularities of these theories and how they can or cannot explain empirical results. Instead, I use the observation of Fehr and Schmidt (1999) that people seem to care about guilt (having more than others) and envy (having less than others) to transition to psychological game theory, which is used in the first manuscript of this dissertation.

In contrast to classic game theory, preferences in psychological game theory depend on both material gains/losses and a person's own or others' beliefs (Battigalli & Dufwenberg, 2022; Geanakoplos et al., 1989). Such models are needed to explain the behavior of participants observed in the first manuscript of this dissertation, where the focus of study lies on the dictator game with an exit option (Dana et al., 2006). The dictator game with an exit option allows participants to avoid playing the game (Lazear et al., 2012) or to avoid the implementation of their allocation decisions (Dana et al., 2006). In Dana et al. (2006), participants were offered a surprise exit option after deciding how to allocate $10 between themselves and an anonymous recipient. If they take the exit option, they get $9, and the recipient will get nothing. Crucially, the recipient will not be informed that he was part of the game if the exit option is taken. The recipient will thus not learn about the actions of the other person. This is crucial, as one explanation of why people take this exit option will involve the beliefs of the potential recipient.

We need psychological game theory to explain this behavior because preferences over monetary outcomes alone cannot explain the behavior of people who take the exit option. To see why, note that the game includes outcomes of ($10, $0) and ($9, $1), and the exit option is thus dominated (in monetary terms) by the options in the dictator game. Thus, the behavior of people who take the exit option cannot be explained by preference models defined solely over (monetary) outcomes (Dana et al., 2006; Krupka & Weber, 2013; Lazear et al., 2012). As typically around half of the participants take the exit option (Cain et al., 2014), distributional preference models certainly seem to miss something. This is where psychological game theory comes into play. While the above-mentioned distributional models also implicitly capture psychological aspects of fairness (e.g., guilt and envy, Fehr & Schmidt, 1999), they still explain behavior solely in terms of preferences over payoffs. In contrast, the model Dana et al. (2006) used to explain exiting in the dictator game is based on the anticipated beliefs of the recipient. Imagine that you were to play the dictator game, and you assume that the recipient expects you to allocate $5 to her. If you were to give her less than $5, you might feel that you let her down by not complying with this assumed expectation. These assumed expectations are called secondorder beliefs, and not meeting them might lead to feelings of guilt (Battigalli & Dufwenberg, 2009). The exit option allows for avoiding these potentially guilt-inducing expectations, as the recipient cannot form expectations if she is not aware of being part of a game. This is aptly captured in the title of the Dana et al. (2006) paper: “What you don’t know won’t hurt me”. Thus, if a participant is sensitive enough to these beliefs, she might accept $9 (instead of $10 when playing the game and giving nothing to the recipient) to keep the recipient in the dark.

Put more formally, Dana et al. (2006) proposed the following utility function:

$$)\equiv X - m - \alpha|\mu - {}_{-}|,$$

where X is the endowment, m is the amount the dictator gives, and μ is the amount that the dictator expects the receiver to expect her to give (i.e., the second-order beliefs). The parameter alpha models the dictator’s sensitivity to the recipient’s expectations (i.e., sensitivity to guilt). Variation in alpha can explain why some participants take the exit option while others don't. If participants exit the dictator game to avoid guilt, we should observe fewer people taking the exit option in a private dictator game treatment where the recipient will not be informed about the game. Indeed, the exit rate dropped to four percent in this treatment by Dana et al. (2006).

The treatment of the dictator game with an exit option was rather extensive because it connects to the overarching theme of this dissertation. While this game is only studied in the first manuscript, all three manuscripts study how prosocial behavior is affected by the context in which it is carried out. In the first manuscript, this is done with the already introduced dictator game with an exit option. This first manuscript aimed to test whether guilt-aversion or self-image concerns can better explain why people avoid the dictator game.

Social Preferences in the Wild

In contrast to the first manuscript of this dissertation, the second and third manuscript study social preferences in a natural environment, i.e., outside the laboratory. While laboratory studies allow for unmatched experimental control, there is debate about the extent to which these laboratory results generalize to naturally occurring situations (Levitt & List, 2007a). While one would hope that observing dictator game allocations in the lab tells us something about how people make donation decisions in the real world, differences in environments can also lead to different observed behavior (Levitt & List, 2007a). Levitt and List (2007a) identify and discuss five factors that can cause differences between behavior observed in the lab and naturally occurring environments: "1) the presence of moral and ethical considerations; 2) the nature and extent of scrutiny of one's actions by others; 3) the context in which the decision is embedded; 4) self-selection of the individuals making the decisions; and 5) the stakes of the game" (p. 154).

We have just seen that self-selection of individuals (point 4) drastically affects dictator game giving in the dictator game with an exit option (Dana et al., 2006; Lazear et al., 2012). Such sorting can naturally occur outside of the lab, and people make use of it to avoid situations that allow/ask for prosocial behavior (Andreoni et al., 2017; DellaVigna et al., 2012). Lacking generalizability from the lab to the field is especially concerning in the domain of social preferences, where practitioners use insights obtained from laboratory studies to boost donations (Bekkers & Wiepking, 2011; Erlandsson, 2021). The second manuscript of this dissertation concerns itself with an effect where there is an incongruity between how the effect is tested for in the lab and the context in which practitioners in the field apply the effect. The effect in question is called the "compassion fade" effect and denotes the phenomenon that subjects in the lab tend to donate less to larger victim groups than to smaller victim groups (Västfjäll et al., 2014). A meta-analysis by Butts et al. (2019) that analyzed 41 studies on this topic has corroborated this finding.

As Garinther et al. (2022) noted, these studies usually present one appeal to each participant so that they see either one person in need or a group of persons, but not both. These so-called separate evaluation designs are problematic, as they "do not adequately reflect the realistic settings in which people make donation decisions" (Butts et al., 2019). A more realistic decision would be to let subjects evaluate multiple donation requests simultaneously (joint evaluation). Because the few existing studies that used a joint evaluation setting found inconsistent results, Garinther et al. (2022) systematically tested how people respond to multiple appeals to help victim groups of different sizes (1, 2, 5, 7, and 12). They found that joint evaluation reverses the compassion fade effect. That is, participants donated more to projects that depicted larger victim groups. Such preference reversals when going from separate to joint evaluations have been observed before (Bazerman et al., 1992), and it has been suggested that people switch from intuitive to more reasoned decision making when going from a separate to a joint evaluation context (Bohnet et al., 2016; Kahneman, 2011). Furthermore, as mentioned by Bohnet et al. (2016), a lack of comparison information in a separate evaluation context has been suggested to lead people to focus on attributes that can most easily be evaluated (Hsee et al. 1999) and rely more on emotional desires (Bazerman et al. 1998). The second manuscript of this thesis examines whether the reversal of the compassion fade effect, as found by Garinther et al. (2022), generalizes to the field. I used data from more than 60’000 crowdfunding projects to test whether crowdfunding projects that depict a larger victim group on their project profile picture attract more donations. Crowdfunding is an ideal context for studying this behavior in a joint evaluation context since people can choose from a myriad of often very similar projects to donate to.

Donating Time Instead of Money

In contrast to the first two manuscripts of this thesis, the third manuscript studies people donating their time. Although volunteering can be modeled as a contribution to a public good (Duncan, 1999) and pure altruism has been suggested to explain volunteering (Unger, 1991), volunteers are also motivated by personal benefits (Govekar & Govekar, 2002). Indeed, the most popular measure to explain volunteering is based on six motives that an individual can fulfill by volunteering, of which only one relates to altruism (Clary et al., 1998). Thus, as is the case for donating money, volunteers' altruistic behavior seems impure. The warm glow people experience from donating their time is even larger than the one they experience from donating their money (Brown et al., 2019; Lilley & Slonim, 2014).

Whether motivated purely or impurely altruistic, volunteers seem to respond to the needs of others (Iizuka &Aldrich; 2022 & Rotolo et al., 2015) and are often willing to risk their health or even lives to help others (Thormar et al., 2016). For example, based on data from volunteering in response to earthquakes, Iizuka and Aldrich (2022) found that the number of deaths and missing persons and the size of the population affected by the disaster correlate most strongly with volunteer turnout. Rotolo et al. (2015) also found a positive relationship between the severity of a crisis and volunteer turnout for US cities affected by the foreclosure crisis. The third manuscript of this dissertation shows that the positive relationship between the severity of a disaster and volunteering could also be observed during the COVID-19 pandemic. Using data from volunteer neighborhood grocery deliveries, we found a concave relationship between the weekly COVID-19 death numbers and the weekly deliveries made by volunteers. We attribute the observed concave effect of the death numbers to the signal of need to help others and the signal of risk to help others conveyed by these numbers.

So, although all three manuscripts study different expressions of prosocial behavior, they all look at how this behavior is influenced by the context it is expressed in. In the first manuscript, the context change is the introduction of the exit option in the dictator game. In the second manuscript, going from a separate to a joint evaluation setting is the change in context. And in the third manuscript, the context is given and changed by the severity of the COVID-19 pandemic.

Manuscript 1: Why do People Avoid Sharing Situations? Testing Guilt Aversion Versus Self-Image Concerns

Meier, D.S., Rieskamp, J., & Schöbel, M. (2022). Why do people avoid sharing situations? Testing guilt aversion versus self-image concerns. Manuscript submitted for publication.

In this manuscript, we tested whether the behavior of people who take the exit option in a dictator game can better be explained by guilt aversion or self-image concerns. In the literature, there is supporting evidence for both motives, but the motives have only been tested in isolation. This is problematic, as the motives could be correlated. If this is the case, a study that only takes into account one of the motives might over- or underestimate the explanatory power of the given motive. We also compared behavior in the dictator game with the behavior in the dictator game with an exit option to see how similar the behavior in the two games is within subjects.

To do this, we let participants play a dictator game and, after at least seven days, a dictator game with an exit option. We also collected the participants' second-order beliefs (guilt) and personal norm ratings (self-image) in the first measurement time point. The personal norm ratings were elicited with the method proposed by Krupka and Weber (2013) to measure injunctive norms, i.e., what one ought to do/what most others approve/disapprove of (Cialdini et al., 1990). In our case, subjects had to rate how appropriate they found a given action in the dictator game with an exit option on a 4-point scale of “very inappropriate” to “very appropriate”. The utility function proposed by Krupka and Weber (2013) assumes that subjects get utility from their own payoff as well as from conforming with the injunctive norm and is able to explain behavior observed in the dictator game with an exit option.

We elicited personal norms instead of injunctive norms because personal norms depend on one's own personal beliefs about what one ought to do and less on socially recognized beliefs about what one ought to do (Burks & Krupka, 2012). While enforcing injunctive norms depends on others, enforcing personal norms does not (Anderson & Dunning, 2014). Given the private setting of the dictator game, we therefore argue that personal norms are more likely to influence behavior in the dictator game than injunctive norms. Because violating one's personal norm results in a negative self-view (Elster, 2007; Fehr & Schurtenberger, 2018), we used conforming to one's personal norm as a proxy for acting in accordance with one's self-image. To operationalize guilt, we elicited participants' second-order beliefs, i.e., what they believe the recipient expects them to give in the dictator game. These beliefs were incentivized. To replicate the results of Dana et al. (2006), in one group, recipients of the dictator game were never informed about the game. This also allowed us to see whether the self-image motive is more important than the guilt aversion motive in such a setting.

Instead of offering participants a simple play/exit option in the dictator game, we elicited their reservation price for playing the game. If someone has a reservation price of $7, they will rather take the exit option and obtain $7 than play the game and divide $10 between them and the recipient. With the reservation price, we thus measure how badly participants want to avoid the game (i.e., how much money they are willing to leave on the table). By letting the reservation price interact with the motives in the mixed logit models that were used to fit the data, we can see how important the motives are in the exit decision.

The results show that guilt (aversion) has a larger effect on an action's utility than selfimage in the dictator game with an exit option. Indeed, self-image only significantly entered the utility function through the interaction with the guilt motive. So, while the coefficients of models that do and do not account for both motives are qualitatively similar to the ones that do, the two motives still interact with each other. The interactions of the motives with the reservation price revealed that the more guilt-averse participants were, the more money they were willing to give up to avoid the game (lower exit reservation price). This is in contrast to the self-image motive, where a higher reservation price (i.e., less avoidance) was associated with a higher utility from a positive self-image. While the results align with previous studies that highlighted the importance of the guilt-aversion motive for the exit decision, they contradict studies that highlighted the explanatory power of injunctive norms. While injunctive and personal norms differ (Bašić & Verrina, 2021), personal norms should be more important in a private setting like the dictator game, as the enforcement of injunctive norms depends on others, while the enforcement of personal norms does not (Anderson & Dunning, 2014). Combined with the results of Krupka and Weber (2013), these results suggest that people who take the exit option think that most others think this is an appropriate thing to do (results of Krupka & Weber, 2013), but they themselves believe it is rather inappropriate (our results).

Manuscript 2: Compassion for all: Real-World Online Donations Contradict Compassion Fade

Meier, D.S. (2023). Compassion for all: Real-world online donations contradict compassion fade. Manuscript under revision.

In the second manuscript, I wanted to test whether the already mentioned reversal of the compassion fade effect in a joint evaluation condition extends to a real-life setting. To test this, I downloaded data from over 60,000 crowdfunding projects from Gofundme.com. Gofundme is the largest donation-based crowdfunding website. $15 billion have been collectively raised since 2010 through projects hosted on this website (GoFundMe, 2022). When browsing projects on Gofundme.com, people see a grid of projects they can donate to. Potential donors are thus in a joint evaluation context when deciding which project to donate to, as they can choose from many projects. Because the compassion fade effect operates via the size of the victim group, I need a way to operationalize the victim group size of the crowdfunding projects. I used the number of persons depicted on a project's profile picture to operationalize the victim group size for two reasons. First, this increases the similarity to stimuli used in lab studies, which typically operationalize victim group size through pictures of people (Garinther et al., 2022). Second, the project profile picture is the dominant visual stimulus that potential donors see when browsing different projects, ensuring that potential donors see the picture before clicking on the project. I used an object recognition algorithm to detect people on the project profile pictures.

Experimental studies can draw causal conclusions about the effect of the victim group size on donations by varying the victim group size of a project while keeping everything else constant (i.e., the cause the money is raised for). We first need to identify this variation from the data to draw causal conclusions with observational data. To achieve this, we need to control for variables that affect both the number of people depicted on the project profile picture and the amount of funds raised by a project. One obvious candidate for such a confounder is the topic of a campaign. Whether the campaign raises money for a sick child or a choir that wants to attend an international event affects both the number of people that will be depicted on the project profile picture and the amount of funds the project will raise. Not controlling for such confounders would lead to spurious associations between the exposure and the outcome. I use the category of a campaign to control for the topic of a campaign. When creating a campaign on

Gofundme.com, one needs to assign the campaign to one of 18 categories (e.g., Medical or Sports). However, stopping here would be similar to an experiment where participants see fundraisers about the same topic but with different descriptions, so still not ideal. I therefore also control for the text that is used to describe the campaign to potential donors.

Controlling for this via a regression model requires translating the text into numbers. I used natural language processing (NLP) methods to do this. These methods encode text into numbers in such a way that similar texts ideally have similar numerical representations. I used three different methods to encode the text into numbers, as this is a critical part of my identification strategy, and different methods have different strengths and weaknesses (Keith et al., 2020). I also control for several other variables that I selected in a principled manner (Wysocki et al., 2022) (e.g., days since the campaign was posted, target amount to be raised, number of times the fundraiser was shared on social media, etc.). Many of these control variables were included to control for the number of people that directly visited the project page of a given fundraiser without browsing through other fundraising campaigns before. Because these people likely only saw one project, they are not in a joint evaluation context, and I thus need to control for this.

The regression results are mostly in line with the findings of a reversed compassion fade effect. While some models do not show an effect, no model shows a significant negative effect of the perceived victim group size (i.e., number of persons on the project profile picture) on the amount of funds raised by a project. For most countries and models, the association is significantly positive for model specifications that are most similar to those of laboratory studies (i.e., only considering projects with a maximum of twelve people on the project profile picture). The results of the robustness checks based on double-machine learning are mostly in line with the results of the regression models. By allowing to control for confounders in a nonlinear way, double machine learning relaxes one of the assumptions of the standard linear regression model (Chernozhukov et al., 2017). I also tested for a quadratic effect in the regression models, which revealed a small but often significant negative quadratic effect.

So, while I was able to replicate the reversal of the compassion fade effect with real-life data, the nonlinearity of the effect might indicate that people might be affected by affective biases even in a joint evaluation setting. According to the affective bias perspective, people's numeracy limitations and biases in affective processing (Hamilton & Sherman, 1996; Slovic, 2007) might be responsible for the compassion fade effect (Butts et al., 2019). Because this effect is quite small, experimental studies might have been underpowered to detect it. In conclusion, this manuscript's results confirm that attributes with a high level of justifiability (i.e., victim group size) trump these biases in a joint evaluation condition. However, the latter still seem to influence decisions to some extent.

Manuscript 3: Risking Your Health to Help Others: Volunteering During the COVID-19 Pandemic

Meier, D.S., Petrig, A., & von Schnurbein, G. (2023). Risking your health to help others: volunteering during the COVID-19 pandemic. *Nonprofit and Voluntary Sector Quarterly.*

The goal of the final manuscript of this dissertation was to test how the severity of the pandemic (i.e., number of cases and deaths) affects the provision of informal volunteering (i.e., grocery food deliveries). In contrast to formal volunteering, informal volunteering takes place outside of the organizational context (Brudney et al., 2019). While there is some evidence from cross-sectional studies that shows that the severity of a catastrophe correlates with volunteer turnout (Iizuka & Aldrich, 2022), there is no evidence on how the severity of a long-lasting catastrophe like the COVID-19 pandemic affects individual helping behavior in the form of informal volunteering. The lack of such evidence makes disaster management suboptimal since professional helping services cannot gauge how informal help will complement their efforts as the catastrophe intensifies or diminishes. Regarding theory building, it is unclear whether the evidence from cross-sectional studies translates to a longitudinal setting.

To test how the severity of the COVID-19 pandemic affects the provision of informal voluntary labor, we used data from volunteer grocery food deliveries. This data was obtained from an app designed to match people who needed groceries delivered (i.e., people who had to self-isolate) with people willing to deliver these groceries voluntarily. This app was launched in March 2020 as a partnership between Migros, Switzerland's largest retailer, and Pro Senectute, a nonprofit organization for the elderly. Until the discontinuation of the app in May 2021 due to low demand, almost 27'000 people registered to do deliveries. These people delivered 72,379 out of 78,961 orders registered on the platform.

Based on theory and the results of previous studies, we hypothesized that weekly case- and death numbers were positively associated with the weekly number of deliveries made by a volunteer. However, since going out to deliver groceries also increased the risk of getting infected with COVID-19, we hypothesized that the association would be concave. A negative quadratic effect of the weekly case- and death numbers on the number of deliveries made would be in line with theories like the health belief model (Janz & Becker, 1984) and the protection motivation theory (Rogers, 1975). We also draw on the appraisal-tendency framework (Lerner &

Keltner, 2000, 2001), which states that fear amplifies risk estimates. These theories suggest that rising case- and death numbers lead to an increase in protective behavior (e.g., staying at home).

We used fixed-effects panel regression models to test the hypotheses as we have multiple observations per individual. Fixed-effects models have the advantage that they control for unobserved factors that are constant over time (e.g., trait prosociality or risk aversion). All models control for such individual fixed effects. Models that used regional (i.e., cantonal) case- and death numbers instead of country-aggregated numbers also control for time fixed effects. Time fixed effects control for time-specific shocks that affected all individuals (e.g., lockdowns). All models also control for the number of orders placed in a volunteer's delivery area and for the number of other volunteers active in the volunteer's delivery area. Because a small group of volunteers made many more deliveries than most other volunteers, we excluded volunteers who made more than 34.3 deliveries (mean + 2 SD). While this only removed 1.4% of the sample (384 volunteers), this small subgroup was responsible for almost half of all deliveries (35,310 deliveries, 48.7%).

The regression results show that the number of deaths, but not the number of cases, positively affected the number of deliveries made across the models. Similarly, the quadratic effect of the death numbers, but not the case numbers, was significantly negative across all models. The same was found for regression models that lagged the case- and death numbers by one week. These lagged models fit the data better, indicating that volunteers need some time to adjust their behavior to a change in the case- and death numbers. So, although the effect of the death numbers on the case numbers is concave, the linear effect dominates the quadratic effect so that the total effect turns negative for very high death numbers. In conclusion, this study replicated cross-sectional evidence on the association between the severity of a catastrophe and the amount of helping behavior. However, in contrast to past studies, we found a concave effect likely caused by the risk associated with the volunteering activity.

General Discussion

By now, social preferences are so established in the behavioral economics literature that the question is no longer whether people are motivated by such preferences but rather why and in which contexts (De Oliveira et al., 2012; Levitt & List, 2007b). In this dissertation, I studied the context dependence of social preferences across three different domains (dictator game giving, giving via crowdfunding, and volunteering). In the first manuscript, we pitched guilt aversion and self-image concerns against each other. We showed that the former could better explain the behavior of people who take the exit option in a dictator game. In the second manuscript, I showed that the reversal of the compassion fade effect in a joint evaluation context extends from the lab to the field (i.e., crowdfunding). And in the third manuscript, we showed that the severity of the COVID-19 pandemic (i.e., the number of deaths) is positively associated with the amount of voluntary labor provided by volunteers. While these manuscripts are related through their field of study (social preferences) and the central research question (context dependence), they build on different pieces of literature. I will therefore discuss how each manuscript builds on and extends this literature separately.

In the first manuscript, we set out to test whether guilt aversion or self-image concerns can better explain people's behavior in the dictator game with an exit option. These motives have so far only been studied in isolation with regard to this question, which is problematic as these motives could be correlated, which would distort the estimated effect of any given motive. Our results showed that in contrast to previous studies that successfully used injunctive norms to explain behavior in this game, personal norms (i.e., self-image) could not explain behavior in the dictator game with an exit option. However, we were able to solidify previous evidence regarding the importance of the guilt aversion motive. Significant interactions between the motives also highlighted the importance of accounting for both motives simultaneously. It is curious to see that injunctive norms (Krupka & Weber, 2013), but not personal norms, can explain behavior in this game. The fact that personal norms were not positively associated with the utility weights of an action in the dictator game with an exit option suggests that participants did not follow their personal norms. The results of the models that were fitted only on participants that did not take the exit option indicate that participants who took the exit option were the ones that did not follow their personal norms.

Maybe people rationalize their personal norms ex-post to justify their egoistic behavior. There is a relatively recent but fast-growing literature showing that people motivate their beliefs in a self-serving manner. These so-called motivated beliefs are especially common in ambiguous situations (Dana et al., 2007). As argued by Gino et al. (2016), judgments of what is moral often possess some flexibility. A motivated Bayesian could exploit this flexibility to pursue egoistic goals while still believing to adhere to moral standards (Gino et al., 2016). This is also in line with recent evidence by Bicchieri et al. (2023). By informing or not informing participants of an upcoming opportunity to lie before eliciting norms regarding lying, they test whether knowledge of the upcoming opportunity to lie affects participants' norms. They found that learning about the upcoming opportunity to lie affects descriptive norms (i.e., empirical expectations of what others will do) but not injunctive norms.

Next to theoretical insights, the first manuscript also provides actionable insights for fundraising. It adds to the literature which shows that people avoid situations in which they are asked to donate money (Andreoni et al., 2017). Also, in line with previous literature, it shows that people are willing to incur costs to do so. To avoid such inefficiencies, fundraisers could give people an easy way to avoid these asks, as too insistent fundraising can also have adverse effects on the organization itself (Adena & Huck, 2020). By providing people with an easy way to avoid the ask, fundraisers can identify the people that are most willing to donate. Kamdar et al. (2015) did this by including a "don't ask me again" option in a mail fundraiser and showed that this condition raised more money than the control condition that did not have such an opt-out option.

Discussing the implications of insights obtained from the lab on real-world behavior naturally brings us to the discussion of the second manuscript of this dissertation. The starting point of the second manuscript was an incongruence between how donation decisions are tested in the lab and how donation decisions are made in the real world. While participants in experiments often only evaluate one project when asked to donate (separate evaluation), in the real world, people often evaluate multiple projects or charities when deciding whom to donate (joint evaluation). Garinther et al. (2022) showed that the compassion fade effect reverses when going from a separate to a joint evaluation context. Using donation data from a domain where people are in a joint evaluation context when deciding whether to donate (crowdfunding), I replicated the results of Garinther et al. (2022) with real-world donation data.

Although the second manuscript is based on observational data, I tried to go beyond correlational analyses. In contrast to economics, where drawing causal conclusions from observational data has a long history, this is still largely taboo in psychological research (Grosz et al., 2020). As argued by Diener et al. (2022), this is unfortunate because a research program is most successful when experiments are not the sole source of valid information but rather are integrated with other methods. This debate also ties back to the discussion mentioned in the introduction about the extent to which laboratory results generalize to naturally occurring situations (Levitt & List, 2007a). Replicating effects that were causally identified in lab experiments allows us to directly address the generalizability of lab results. As argued by Harrison and List (2004), combining lab experiments with field data allows for sharper and more convincing inferences. As highlighted by my discovery of a concave relationship between the number of people on a project profile picture (i.e., perceived victim group size) and the amount of funds raised by a project, field data can also reveal phenomena that lab experiments have missed. In this case, lab studies might have been underpowered to detect this effect because the nonlinear effect is small in magnitude.

In the third manuscript, we also identified a concave relationship but between the weekly COVID-19-related death numbers and the amount of voluntary labor provided by volunteers. We were able to replicate previous cross-sectional evidence of the positive association between the severity of a catastrophe and the amount of volunteering in a longitudinal sample. But in contrast to previous results, the association we found was concave. We attribute this to the risk entailed with volunteering during the COVID-19 pandemic. While this is plausible, we are limited in attributing observed behavior to specific motives due to the study's observational nature. This is especially the case given that there are many motives that can explain volunteering above and beyond those used to explain giving money (Clary et al., 1998). The proposed negative effect of the risk of volunteering on the provided amount of voluntary work connects to laboratory studies that show that participants donate less when there is a greater risk that their donation will have less impact (Brock et al., 2013). Tying back to the first manuscript of this dissertation, Exley (2016) showed that people exploit risk in a self-serving manner. She showed that when participants faced tradeoffs between money for themselves and the charity, they acted more averse to charity risk and less averse to self-risk. While I don't think that the risk of volunteering was actively used as an excuse not to volunteer in the context of the third study of this dissertation, future research could investigate whether volunteers also exploit excuses or engage in motivated reasoning to avoid volunteering opportunities.

In conclusion, this dissertation adds to the literature by testing how prosocial behavior is affected by changes in the context in which it is carried out. In the first manuscript, we pitched two motives against each other that can potentially explain the context-induced behavior change (i.e., adding an exit option to the dictator game). In the second manuscript, I showed that the reversal of the compassion fade effect when going from a separate to a joint evaluation context extends to online giving behavior (i.e., crowdfunding). And in the third manuscript, we showed that the severity of a catastrophe (i.e., the COVID-19 pandemic) affects the amount of catastrophe-related voluntary labor provided during the catastrophe (i.e., the COVID-19 pandemic).

References

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization, 3*(4), 367-388.

Abrams, B. A., & Schitz, M. D. (1978). The'crowding-out'effect of governmental transfers on private charitable contributions. *Public Choice*, 29-39.

Adena, M., & Huck, S. (2020). Online fundraising, self-image, and the long-term impact of ask avoidance. *Management Science, 66*(2), 722-743.

Anderson, J. E., & Dunning, D. (2014). Behavioral norms: Variants and their identification. *Social and Personality Psychology Compass, 8*(12), 721-738.

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., Rao, J. M., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of political Economy, 125*(3), 625-653.

Bašić, Z., & Verrina, E. (2021). Personal norms—and not only social norms—shape economic behavior. *MPI Collective Goods Discussion Paper*, (2020/25).

Battigalli, P., & Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory, 144*(1), 1-35.

Battigalli, P., & Dufwenberg, M. (2022). Belief-dependent motivations and psychological game theory. *Journal of Economic Literature, 60*(3), 833-882.

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative science quarterly*, 220-240.

Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of management review, 23*(2), 225-241.

Becker, G. S. (1974). A theory of social interactions. *Journal of political economy*, *82*(6), 1063-1093.

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.

Bicchieri, C., Dimant, E., & Sonderegger, S. (2023). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior, 138*, 321-354.

Bohnet, I., Van Geen, A., & Bazerman, M. (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science, 62*(5), 1225-1234.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review, 91*(1), 166-193.

Brock, J. M., Lange, A., & Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, *103*(1), 415-437.

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.

Brudney, J. L., Handy, F., & Meijs, L. C. (2019). Introduction to the special issue: Exploring the dimensions of volunteering. *Nonprofit and Voluntary Sector Quarterly, 48*(2_suppl), 5S-11S.

Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science, 58*(1), 203-217.

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes, 151*, 16-33.

Cain, D. M., Dana, J., & Newman, G. E. (2014). Giving versus giving in. *The Academy of Management Annals, 8*(1), 505-533.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review, 107*(5), 261-265.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology, 58*(6), 1015.

Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology, 74*(6), 1516.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes, 100*(2), 193-201.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 67-80.

Daruvala, D. (2010). Would the right social preference model please stand up!. *Journal of Economic Behavior & Organization, 73*(2), 199-208.

De Oliveira, A., Croson, R. T., & Eckel, C. C. (2012). Are preferences stable across domains? An experimental investigation of social preferences in the field. *Southern Economic Journal, 79*(1).

De Wit, A., & Bekkers, R. (2017). Government support and charitable donations: A metaanalysis of the crowding-out hypothesis. *Journal of Public Administration Research and Theory, 27*(2), 301-319.

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics, 127*(1), 1-56.

Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science, 17*(4), 1101-1119.

Duncan, B. (1999). Modeling charitable contributions of time and money. *Journal of public economics, 72*(2), 213-242.

Elster, J. (2015). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge University Press.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics, 14*, 583-610.

Erlandsson, A. (2021). Seven (weak and strong) helping effects systematically tested in separate evaluation, joint evaluation and forced choice. *Judgment and Decision Making, 16*(5), 1113-1154.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of*

Economic Studies, *83*(2), 587-628.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics, 114*(3), 817-868.

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature human behaviour, 2*(7), 458-468.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior, 6*(3), 347-369.

Garinther, A., Arrow, H., & Razavi, P. (2022). Victim number effects in charitable giving: Joint evaluations promote egalitarian decisions. *Personality and Social Psychology Bulletin, 48*(1), 34-48.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and economic Behavior*, *1*(1), 60-79.

Gino, F., Norton, M. I., & Weber, R. A. (2016). Motivated Bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives, 30*(3), 189-212.

GoFundMe (2022, July 7). *GoFundMe's 2021 giving report.* https://www.gofundme.com/c/gofundme-giving-report-2021

Govekar, P. L., & Govekar, M. A. (2002). Using economic theory and research to better understand volunteer behavior. *Nonprofit Management and Leadership, 13*(1), 33-48.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science, 15*(5), 1243-1255.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization, 3*(4), 367-388.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological review, 103*(2), 336.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature, 42*(4), 1009-1055.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review, 91*(2), 73-78.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological bulletin, 125*(5), 576.

Iizuka, A., & Aldrich, D. P. (2022). Attracting altruists: explaining volunteer turnout during natural hazard events in Japan. *Disasters, 46*(2), 526-544.

Janz, N. K., & Becker, M. H. (1984). The health belief model: A decade later. *Health education quarterly, 11*(1), 1-47.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business, 59*(4), 285-300.

Kamdar, A., Levitt, S. D., List, J. A., Mullaney, B., & Syverson, C. (2015). Once and Done: Leveraging Behavioral Economics to Increase Charitable Contributions.". In *SPI Working Paper*.

Keith, K. A., Jensen, D., & O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. arXiv preprint arXiv:2005.00649.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association, 11*(3), 495-524.

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics, 4*(1), 136-163.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & emotion, 14*(4), 473-493.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology, 81*(1), 146.

Levitt, S. D., & List, J. A. (2007a). On the generalizability of lab behaviour to the field. *Canadian Journal of Economics/Revue canadienne d'économique, 40*(2), 347-370.

Levitt, S. D., & List, J. A. (2007b). What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic perspectives, 21*(2), 153174.

Lilley, A., & Slonim, R. (2014). The price of warm glow. *Journal of Public Economics, 114*, 58-74.

Ottoni-Wilhelm, M., Vesterlund, L., & Xie, H. (2017). Why do people give? Testing pure and impure altruism. *American Economic Review, 107*(11), 3617-3633.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude chang. *The journal of psychology, 91*(1), 93-114.

Rotolo, T., Wilson, J., & Dietz, N. (2015). Volunteering in the United States in the aftermath of the foreclosure crisis. *Nonprofit and Voluntary Sector Quarterly, 44*(5), 924-944.

Slovic, P. (2007). “If I look at the mass I will never act”: Psychic numbing and genocide. *Judgment and Decision making, 2*(2), 79-95.

Thormar, S. B., Sijbrandij, M., Gersons, B. P., Van de Schoot, R., Juen, B., Karlsson, T., & Olff, M. (2016). PTSD symptom trajectories in disaster volunteers: The role of self-efficacy, social acknowledgement, and tasks carried out. *Journal of Traumatic Stress, 29*(1), 17-25.

Unger, L. S. (1991). Altruism as a motivation to volunteer. *Journal of economic psychology, 12*(1), 71-100.

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PloS one, 9*(6), e100115.

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science, 5*(2), 25152459221095823.

Why do people avoid sharing situations? Testing guilt aversion versus self-image concerns

Dominik S. Meier[1, 2, *], J¨org Rieskamp[1], and Markus Sch¨obel[1]

[1]University of Basel, Department of Psychology, Missionsstrasse 62a, 4055 Basel, Switzerland
[2]University of Basel, Center for Philanthropy Studies, Steinengraben 22, 4051 Basel, Switzerland
[*]Corresponding author: Dominik S. Meier, dominik.meier@unibas.ch

June 26, 2023

Abstract

People sometimes avoid situations that allow for sharing with others even if this behavior involves some costs. Because social preference theory cannot explain this behavior, the concepts of guilt aversion and self-image concerns have been proposed as motivational explanations. However, current evidence does allow to conclude which concept is more important. To test both concepts rigorously against each other, we conduct an experimental study using dictator games with and without an exit option. We operationalized guilt by eliciting second-order beliefs and self-image concerns by eliciting personal norms. The study used a between-subjects design, wherein one condition the recipients in the dictator game were not informed about the type of game they were playing, thereby ruling out guilt aversion as a driving motivation for the dictators and leaving selfimage concerns as the probable explanation for taking the exit option. The results show that whereas dictators in the informed condition, where guilt was possible by design, were willing to give up more money to avoid the game, they on average also gave more conditional on playing the game than dictators in the not-informed condition. But due to the increase in avoidance, guilt aversion did not increase aggregate giving. Guilt aversion thus both drives people out of sharing situations but also propels them to give once the situation can no longer be avoided. Mixed logit models also suggest that people avoid sharing situations to avoid guilt and not to preserve their self-image.

Keywords: dictator game; exit option; personal norms; self-image; guilt
Declarations of interest: none

1 Introduction

Past research illustrating that people share resources with others and follow fairness principles challenged the view of purely selfish human beings. However, more recent work has shown that people's apparent prosocial behavior is less social than previously thought. A variety of recent field and lab experiments have shown that people are willing to pay a cost to avoid situations that allow for prosocial behavior. For instance, Andreoni et al. (2017) observed that some people took a detour to avoid solicitors that were positioned at one or both of the entrances of a mall. In cold weather, however, this was no longer observed (Trachtman et al., 2015), showing that people are sensitive to the costs of avoiding a situation that allows for sharing (hereafter also referred to as a "sharing situation"). A field experiment by DellaVigna et al. (2012) provided evidence that people also avoid solicitors when they are at home: Residents who were informed with a flyer about the time solicitors would stop by answered the door less often than residents who did not get this heads-up. This suggests that residents who got the heads-up made sure that they were not at home when the solicitors planned to visit. Adena and Huck (2019) provided evidence that such avoidance can also have an adverse long-term impact on businesses that partner with charities for fundraising. They found that opera customers who faced more insistent online fundraising bought fewer opera tickets in the next season. People even avoided recycling machines that allowed them to donate the returned deposit (Knutsson et al., 2013). The present work aims to explain why people avoid sharing situations.

Lab experiments that allowed participants to avoid a sharing situation also found that avoidance is widespread, and that egoism is far more prevalent than the results from studies that do not allow for avoidance suggest. Many of these studies used the dictator game, where two people are paired and one of them, called the "dictator", gets an endowment that they may split with the other person, called the recipient. The recipient has no say in the game, hence the game's name. Avoidance of this game is made possible by adding an "exit" option. This allows dictators to avoid the game (Lazear et al., 2012) or to prevent the implementation of the game (Dana et al., 2006). Dana et al. (2006) o↵ered dictators the exit option after they decided how to allocate $10 in the dictator game, but before the game had been implemented and before the recipients were informed about the game. Dictators were o↵ered $9 if they chose to exit the game. If the dictator chose to exit, the recipient was not informed about the game (i.e., they did not know that they were part of the game) and remained empty-handed. In their first study, almost one-third of the participants chose to exit (Dana et al., 2006). This resulted in a stark increase in egoistic behavior since almost all of those dictators had initially decided to give something to the recipient. Lazear et al. (2012) reported a similar detrimental e↵ect of the option to avoid the game on aggregate giving. They estimated that one-third of the participants would share the endowment when the dictator game could not be avoided but would avoid the game and hence share nothing when the exit option was available. Lazear et al. (2012) called this phenomenon *reluctant sharing*.

Reluctant sharers share when they are in a situation that cannot be avoided and that allows for sharing. However, when given the opportunity, they avoid the sharing situation or are even willing to incur a cost to revoke the sharing decision they made when avoidance was not possible. Broberg et al. (2007) showed that the costs of being in a sharing situation where exiting is not possible are high. To exit the game, participants were on average willing to forego 18% of the dictator game endowment. In sum, studies from the field and from the lab show that a significant proportion of people prefer to avoid sharing situations and give nothing but feel compelled to give once they face the sharing situation and have the possibility to give money. Cain et al. (2014) referred to this behavior as "giving in." Cain et al. (2014) estimated that up to half of what has traditionally been thought of as unconditional giving (in lab studies) is giving in.

The notion of giving in implies that people sometimes fail to follow their true underlying preferences when making a decision. This implication is reinforced by observations that people who exit the dictator game neither comply with the selfishness axiom (when they exit the game at cost) nor seem to truly care about the welfare of the recipient. A pure egoist should play the game and keep the whole endowment (assuming the exit option yields a lower endowment than the dictator game), whereas someone who cares about the recipient should play the game and give something to the recipient. The exit option should not be chosen since it is a dominated option in this simple egoism–altruism framework. Indeed, distributional models of social preferences, which describe preferences in terms of the distribution of payoffs to the involved parties, cannot explain this behavior (see the online appendix of Krupka and Weber (2013) for a detailed account). Therefore, understanding why people avoid sharing situations also furthers an understanding of why people give in the first place. Although avoidance of sharing situations is widespread and can have negative long-term consequences (Adena and Huck, 2019), the causes of this behavior are not well understood. Our study was intended to close this gap.

In the following, we briefly discuss the two motives that have been commonly suggested to explain why people take the exit option (see section 2 for a detailed account): guilt aversion and self-image concerns. Both motives can explain why people avoid situations that allow for sharing and give only if the situation cannot be escaped. In general, the concept of guilt aversion implies that people share because they want to avoid feeling guilty (Dana et al., 2006). In behavioral economics, guilt is often modeled based on awareness of another person's expectations (Battigalli and Dufwenberg, 2007). Not meeting those expectations may lead to feelings of guilt. The exit option offers a way to avoid these expectations because in this case, the recipient is not even aware that they could have been part of a dictator game. Thus, by exiting, the dictator can avoid feelings of guilt, since the prerequisites for experiencing guilt are not met when the exit option is chosen. However, if an exit option is not available, the recipient is informed about the game, and the prerequisites for experiencing guilt are thus met. This may lead people to give to avoid feeling guilty. In line with this explanation, Dana et al. (2006) found that when dictators were aware that recipients in the dictator game would not be informed about the game at all (irrespective of whether the dictator played the game or exited), significantly fewer of them took the exit option, presumably because now the exit option lost its appeal because the prerequisites for experiencing guilt were also not met if dictators decided to play the game.

According to the second explanation, people take the exit option because it offers a relatively cheap way of maintaining one's self-image of being a moral person. Following a Bayesian signaling model, Grossman and van der Weele (2017) conjectured that "the people who opt out of dictator games may have initially contributed mainly because of image concerns" (p. 206). This explanation suggests that people consider exiting the dictator game as more morally appropriate than playing the game and giving very little. Krupka and Weber (2013) provided supporting evidence by focusing on the influence of injunctive norms (i.e., focusing on what people believe others expect from them) on exiting. They showed that taking the exit option is seen as more socially appropriate than playing the dictator game and keeping most of the endowment. This makes the exit option very attractive for people who care about their moral self-image because exiting yields a similar payoff to playing the game and keeping most of the endowment but eliminates a potential threat to the self-image. A model based on the social appropriateness and the payoff of a given action provided a good explanation of the data from the dictator game with an exit option (Krupka and Weber, 2013).

It can be seen from the above that both motives can explain sharing behavior in standard dictator games and can also explain the avoidance of sharing situations when an exit option is available. However, the explanatory power of these motives has not been tested in combination. This is an important shortcoming since these motives might be correlated, meaning that people that are more guilt averse are also more susceptible to self-image concerns. If this is the case, the explanatory power of a single motive could be overstated if one does not control for the other motive. In this study, we set out to address this issue by elicitating both motives. This allows us to test which of the two motives, self-image concerns or guilt aversion, has more explanatory power in regard to the behavior in the dictator game and the dictator game with an exit option.

The remainder of this paper is organized as follows: Section 2 presents the theory and Section 3 the experimental design. The results are presented and discussed in Section 4, and Section 5 concludes.

2 Theory

2.1 Guilt aversion

Several studies have shown that some people are sensitive to the recipient's expectations in economic games such as the dictator game (Balafoutas and Fornwagner, 2017; Balafoutas and Sutter, 2017). This finding has also been confirmed in a fundraising context (Edwards and List, 2014). In the framework of behavioral economics, not meeting those expectations leads to feelings of guilt (Battigalli and Dufwenberg, 2007; Charness and Dufwenberg, 2006; Hauge, 2016). To avoid feeling guilty, a guilt averse decision maker tries to live up to others' expectations. This operationalization, which relies solely on expectations, differs from those used in other disciplines (e.g., in psychology; Tangney et al., 2007). Although defining guilt by relying solely on the expectations of others might seem narrow, it allows for parsimonious and testable models. In the dictator game, guilt (aversion) usually drives people to give more. However, if an exit option is present, guilt aversion can have the opposite effect and drive people out of the sharing situation (Dana et al., 2006). Within the framework of psychological game theory, Dana et al. (2006) used guilt aversion to explain why participants took the exit option. In contrast to classic game theory, preferences in psychological game theory depend on both material gains/losses and a person's own or others' beliefs (Battigalli and Dufwenberg, 2020). Dana et al. (2006) adapted Battigalli and Dufwenberg's (2009) utility model such that the utility of a given option for the dictator depends not only on the payoff of the option but also on the perceived expectations of the recipient. The utility of a given action a_k is then given by:

$$u(a_k) = \alpha\pi(a_k) + \gamma\max(\mu - m, 0), \tag{1}$$

where $\pi(a_k)$ is the monetary payoff produced by the selected action a_k, α is the value the individual places on this payoff, μ is the amount the dictator thinks the recipient expects the dictator to give (second-order beliefs), and m is the amount the dictator gives. The parameter γ is used to represent the individual dictator's sensitivity to the receiver's expectations. If $\gamma < 0$, the dictator experiences disutility from not meeting the recipient's expectations and is thus guilt averse. The weights α and γ can be estimated using conditional logit (McFadden, 1973) on equation (1). Following this model, the dictator will adjust their behavior to the recipient's expectations if γ is not 0. When the dictator is guilt averse, the utility will be higher if the recipient has no expectations of receiving something, since this will eliminate the term weighted by γ (which in this case would be negative). And this is exactly what the exit option offers: Recipients will not be informed about the game if the dictator exits and thus cannot form expectations about how much money they will get.

According to Dana et al. (2006), dictators choose the exit option to avoid being mentally confronted with the recipient's expectations. In contrast to our design, which is based on Lazear et al. (2012), dictators in Dana et al. (2006) faced a surprise exit option that was introduced only after they had already played the dictator game, but before it was implemented. In contrast, in our study dictators faced an upfront exit option, allowing them to completely forego playing the game. Whether participants face an upfront or surprise exit option should have no influence on the effect of guilt (aversion). We opted to use this design since it has not yet been tested with regard to the guilt motive and since it is more reflective of real-life situations. To test the guilt aversion model, Dana et al. (2006) conducted a second study. They added a *private* condition where the recipients were never informed about the game and thus were unable to form expectations. As predicted by the model, significantly fewer dictators (who were aware the recipients were not informed) exited in this “not-informed” condition. This indicates that dictators may feel guilty about not giving if the recipient is able to form expectations. This evidence suggests that one reason dictators take the exit option is to avoid guilt. We call this the guilt aversion motive.

2.2 Self-Image Concerns

Some authors have argued that self-image concerns can explain why some dictators take the exit option (Grossman and van der Weele, 2017; Tonin and Vlassopoulos, 2013). Having moral traits is considered the most important part of one’s self-image, as shown by Strohminger and Nichols (2014). In the words of the authors: “The self is not so much the sum of cognitive faculties as it is an expression of moral sensibility; remove its foothold on that world, and watch the person disappear with it” (p. 169). For self-image concerns to explain why people take the exit option, taking the exit option should be deemed more appropriate than playing the game and keeping the whole endowment. If this is the case, taking the exit option results in a higher self-image than playing the game and keeping the whole endowment, but the payoff of both actions is the same. Indeed, a study by Krupka and Weber (2013) showed that taking the exit option was more socially appropriate than playing the game and keeping less than $3 (given a $10 endowment). Krupka and Weber (2013) elicited the injunctive social norm, that is, what one ought to do/what most others approve/disapprove of (Cialdini et al., 1990). They did this by letting participants rate the social appropriateness of all possible even-dollar allocations in a dictator game on a 4point scale of “very socially inappropriate” to “very socially appropriate.” The task was designed as a coordination game: When the participant’s rating of a randomly chosen allocation corresponded to the modal response, the participant got a reward, ensuring the incentive compatibility of the elicitation mechanism. In addition, participants also rated the option to exit the game. In contrast to Krupka and Weber (2013), we used personal norms instead of injunctive norms as another means to explain why the dictator game is avoided. Personal norms differ from injunctive norms in that they depend less on socially recognized beliefs about what one ought to do and more on one’s own *personal* belief about what one ought to do (Burks and Krupka, 2012). While the enforcement of injunctive norms depends on others, the enforcement of personal norms does not (Anderson and Dunning, 2014). Given the private setting of the dictator game, we therefore argue that personal norms are more likely to influence behavior in the dictator game than injunctive norms. Schwartz (1977) defined personal norms as “self-expectations for specific action in particular situations that are constructed by the individual” (p. 227). According to Schwartz (1977), personal norms create feelings of moral obligation. Since violation of personal norms results in a negative self-view (Elster, 2007; Fehr and

Schurtenberger, 2018), we used conforming to one's personal norm as a proxy for acting in accordance with one's self-image.

For eliciting personal norms we used the same method as Krupka and Weber (2013), but instead of asking for the participants' social appropriateness rating (i.e., what they thought most others think is the appropriate thing to do), we elicited how personally appropriate participants thought each action to be. That is, participants evaluated all possible actions according to their own *personal* norms. This was done without incentivization since one cannot easily incentivize personal beliefs.

Krupka and Weber (2013) used their norm data to explain the results of Lazear et al. (2012). The utility function proposed by Krupka and Weber (2013) assumes that subjects get utility from their own payoff, $\pi(a_k)$ as well as from conforming with the injunctive norm, $N(a_k)$, where 0 is the weight given to adhering to the injunctive social norm of action k, $N(a_k)$,

$$u(a_k) = V(\pi(a_k)) + N(a_k).$$

Krupka and Weber's (2013) results show that the injunctive social norm was more in favor of taking the exit option than playing the game and sharing less than \$3. Therefore, an individual who cares about adhering to social norms (0) and would share less than \$3 in the dictator game can increase their utility by taking the exit option. By imposing a linear restriction on $V() : V(\pi) = \pi$, the weights given to monetary payoff and social appropriateness can be estimated using conditional logit on the following equation:

$$u(a_k) = \pi(a_k) + N(a_k). \quad (2)$$

As explained above, we used personal instead of injunctive norms, and we used the parameter as a proxy for self-image. Why can this utility function explain taking the exit option? Because the exit option yields a high payoff while at the same time being relatively socially appropriate, allowing participants to maintain their self-image (Krupka and Weber, 2013). The results of Krupka and Weber (2013) thus imply that the exit option is especially attractive for people who care about their payoff as well as their self-image.

2.3 Combining the motives

As shown above, both guilt aversion (Dana et al., 2006) and self-image concerns (Grossman and van der Weele, 2017; Krupka and Weber, 2013) have been proposed to explain exiting the dictator game. Evidence from the private condition of Dana et al. (2006) suggests that guilt aversion is a powerful motive since almost no one exited when the dictator game was private from the beginning. However, there are also studies that have shown that exiting occurs when guilt is not at play, leaving self-image concerns as the likely explanation (Tonin and Vlassopoulos, 2013) Thus, although both motives have supporting evidence, the motives have never been tested in combination. As already mentioned, this is a shortcoming since the two motives could be correlated which would result in a possible over- or understatement of the considered motive at the cost of the not considered motive. The primary goal of this study was to find out which of these two motives, guilt aversion or self-image concerns, has a higher explanatory power in the decision to play or exit a dictator game. By combining the already introduced utility functions and fitting them with conditional logit models (McFadden, 1973), we can test whether guilt aversion or self-image concerns matter more in the decision to play or exit the game:

$$u(a_k) = \uparrow(a_k) + \leftarrow \max(\mu \quad m, 0) + N(a_k) \tag{3}$$

2.4 Hypotheses

Instead of giving participants a simple choice of playing or not playing the dictator game (Lazear et al., 2012), we measured dictators' valuation of playing the game with the multiple price list (MPL) mechanism (Andersen et al., 2006). Broberg et al. (2007) did this for the sequential exiting paradigm (Dana et al., 2006) and found that dictators on average were willing to give up 18% of their endowment to avoid the implementation of the dictator game. This method had not been used yet for the upfront exiting design introduced by Lazear et al. (2012). This method makes it possible to measure how strongly dictators want to play or exit the game and to relate this to guilt aversion and self-image concerns. It is thus a more nuanced measure than a simple play/exit decision. The lower the exit reservation price determined with this method, the lower the probability that the game will have to be played.

Our first hypothesis addresses the effect of guilt aversion of a person in the dictator role when a recipient's expectations cannot be escaped, that is, when there is no exit option. In this case, the only way a guilt averse dictator can avoid feeling guilty is by meeting their second-order beliefs about the recipient's expectations. Since second-order beliefs are by design only possible in the informed condition, we should see higher aggregate giving in this condition if dictators are guilt averse.

Hypothesis 1: In the standard dictator game, aggregate giving is higher in the informed condition than in the not-informed condition.

It is important to note that this and the following hypothesis are not comparing the explanatory power of guilt-aversion and self-image concerns. It is rather a replication of the findings of Dana et al. (2006) with a dictator game with an exit option that is 1) implemented with a multiple price list and 2) introduced ex-ante and not ex-post.

If an exit option is available, dictators can exit the game to avoid feeling guilty, because in this case the recipient is not informed about the game and there are thus no expectations one could not meet. Our second hypothesis addresses the theory that people take the exit option to avoid guilt (Dana et al., 2006). Following

Dana et al. (2006), we also implemented an informed and a not-informed condition (see section 3 for more details). Because guilt is ruled out by design in the not-informed condition, this benefit of the exit option is unique to the informed condition. We therefore expect that participants in the informed condition are willing to give up more money to take the exit option than participants in the not-informed condition.

Hypothesis 2: The average reservation price for exiting the game is lower in the informed than in the not-informed condition (i.e., people are willing to pay more to exit the game in the informed than in the not-informed condition).

The next two hypotheses allow us to test which motive has a higher explanatory power with regard to taking the exit option. To do this, we make use of the exit reservation price. The coe cients of model (3) will tell us how important a motive is in explaining behavior in the dictator game with an exit option as a whole, not just with regard to the decision of whether to exit the game. Because the exit reservation price influences the probability that one has to play the game (i.e., the lower the reservation price for exiting the game, the less likely one is to play the game), we can test whether these motives influence the decision to exit the game by letting them interact with the exit reservation price. If these motives lead people to take the exit option, the e↵ect of these motives on the behavior in the game should be strongest for people who want to exit the game (i.e., people with a low reservation price). The interaction allows us to test this hypothesis. If guilt aversion drives people to take the exit option, we expect that the lower the exit reservation price, the stronger the e↵ect of guilt aversion is (i.e., the lower ↵/ is). Thus for people with a low exit reservation price, the e↵ect of guilt aversion on the utility of an action should be stronger than for people with a high exit reservation price. Note that we should only observe this in the informed condition since the prerequisites for experiencing guilt are by design not met in the not-informed condition.

Hypothesis 3: There is a positive interaction between guilt aversion and the exit reservation price in the informed condition.

We also use the exit reservation price to test how important the self-image motive is for taking the exit option. As stated above, we argue that self-image enforces the norm-following behavior in a private setting such as ours (anonymous online dictator game). Considering that in Krupka and Weber (2013) taking the exit option was rated more socially appropriate than playing the game and giving very little, we should see that the more money participants are willing to forego to exit the game (i.e., the lower the reservation price), the more utility they should get from following their personal norms and thus maintaining a positive self-image. Given our setting, this means that we expect that the lower the exit reservation price, the stronger the e↵ect of self-image is (i.e., the higher is). Thus for people with a low exit reservation price, the utility weight of self-image should be larger than for people with a high exit reservation price.

Hypothesis 4: There is a negative interaction between self-image and the exit reservation price.

Although our methodology draws heavily from behavioral economics, the measured motives originated in psychology. We therefore also used a method that is common in psychology to measure those two motives, the Test of SelfConscious Affect–3 (TOSCA-3) guilt and shame proneness questionnaire (Tangney et al., 2000). This measure is more trait-based than our other measures since it measures general guilt and shame proneness (Tangney et al., 1992). Shame is very closely associated with self-image since it arises when people evaluate a

Procedure
T1
DictatorGameInstructions&ComprehensionQuestionwithFeedback
DictatorGame
Second-orderbeliefselicitation
Demographicinformationcollected
Instructionsonpersonalnormratings,comprehensionquestionswithfeedback,personalnormartings
TOSCA-3questionnaire
Breakofsevendays
T2
DictatorGamewithMPLexitoption
Payment(Oneofthetwogamesrandomlychosenforbonuspayment+bonusfromsecond-orderbeliefs)

Table 1: Procedure of the experiment in chronological order.

threat to their self (Daniels and Robinson, 2019). The definition of guilt as used in behavioral economics is very specific in that it relies solely on second-order beliefs. Psychologists traditionally consider guilt in a broader sense (Tangney et al., 2007) and use questionnaires such as the TOSCA-3 to measure guilt proneness. A study by Bellemare et al. (2019) showed that the economic method of measuring guilt (second-order beliefs) and the psychological method (TOSCA-3 in this example) correlate at about 0.3. Previous research that used the TOSCA3 questionnaire found that guilt proneness, but not shame proneness, correlated with prosocial behavior (Bracht and Regner, 2013). Including the TOSCA-3 guilt and shame measures in the mixed-logit models gave us a second way of testing whether guilt aversion or self-image concerns matter more in the decision to take the exit option. The next section describes the measures in more detail.

3 Methods

3.1 Experimental Design

We used a between-subjects design with two conditions (informed and notinformed) and two measurement time points (T1 and T2) to answer our research question. The only difference between the two conditions was that in the not-informed condition, recipients were not informed about the dictator game, regardless of whether dictators played the game, and dictators were aware of this stipulation. Specifically, in the not-informed condition, dictators were told: “Importantly, Individual B will never find out about the game and about how much money (points) you decided to give to him/her. Thus, in case that you decided to give something to Individual B, Individual B will only be informed that the money (points) comes from an experiment on Prolific, without further details.” In contrast, dictators in the informed condition read the following: “Importantly, Individual B will be informed about his/her role (as recipient) in the game and about how much money (how many points) you decided to give to him/her.” A similar di↵erence in instructions occurred in the personal norm elicitation and the dictator game with an exit option. Apart from this, the procedure in both conditions was the same. The full instructions can be found in the Appendix [2]. No deception was used at any time. We used two measurement time points to avoid spillovers from the belief elicitations on the behavior in the dictator game with an exit option. We chose to separate the two measurement time points by at least one week because this is in the range of what has been done in similar studies (G¨odker et al., 2021; Amelio and Zimmermann, 2023).

3.2 Procedures

Participants playing the role of the dictator were randomly assigned to one of the two between-subjects conditions. The only di↵erence between the conditions was the instructions regarding whether the recipient was informed or not, as mentioned above. Thus the procedure described here was the same for both conditions. At T1, participants first played a dictator game with an endowment of £5, which could be allocated in the form of 10 points. After participants read the instructions but before they played the game, we asked them a comprehension question regarding the between-subjects manipulation. [3] After answering the comprehension question, we gave participants feedback to make sure that all participants understood the design. Participants then played the dictator game. After this, second-order beliefs were elicited. This was done in an incentivized way: dictators got a bonus of £0.5 if they correctly predicted the expectation of the recipient. Since recipients in the not-informed condition were not informed about the game, eliciting second-order beliefs in this way would not make sense. To circumvent this, we told participants in the not-informed condition to predict what they thought an informed recipient would expect in the game.

Participants then filled in demographic information (enrolled in university, age, sex). After this, we collected participants’ personal norms regarding the

[2] https://osf.io/hx2kf/?view_only=27ac2bf64d1d4ef992071382bfa8b118

[3] Participants had to state whether the statement “Individual B will be informed about the game and about how much I decided to allocate to him/her” was true or false. In both conditions, 95% of participants answered the comprehension question correctly. Participants were aware that we would ask a comprehension question. We expected that this would lead participants to pay more attention to the instructions.

dictator game with an exit option. Following the mechanism proposed by Krupka and Weber (2013), participants rated the appropriateness of all options available in the game. Specifically, participants were instructed to "base this rating on what you think is the right thing to do according to your personal opinion." They could do this on a 4-point scale: "very inappropriate" [1], "inappropriate" [1/3], "appropriate" [1/3], and "very appropriate" [1]. The corresponding numerical values used in the analysis (shown in the brackets) were not shown to the participants. Before participants completed this task, we asked two comprehension questions in the same way as in the dictator game.[4] As the last task of part 1, participants filled in the TOSCA-3 questionnaire.

After a break of at least seven days, participants could start with the second part of the study. We did this to avoid spillover e↲ects from the belief elicitations and from the dictator game on the behavior in the dictator game with an exit option at T2. In the second part, participants played a dictator game with a MPL (Andersen et al., 2006) exit option. The multiple price list was implemented as follows: Participants playing the role of the dictator were presented with 24 rows each containing the choice of (A) to play the game or (B) to opt out of the game for x points, where x = 0.5 * row number. Participants only had to indicate their switching point (i.e., at which value of x they preferred x points for exiting the game to playing the game). Consistency was enforced (i.e., multiple switching points were not possible). The switching point is equivalent to the reservation price for taking the exit option. After participants revealed their switching point, one row was randomly chosen and the preferred alternative in this row was implemented. Thus, the lower the switching point (reservation price), the less likely it was that the game had to be played. If playing the game was the preferred option in the randomly chosen row, participants played a dictator game just as they did in part 1. If opting out was the preferred option, participants received 0.5 * chosen row number points and did not play the game. Participants playing the role of the recipient were informed about the game only if the game had been played in the informed condition. Participants were aware from the beginning that only one of the two dictator games would be randomly implemented and determine their bonus payment.

Participants in the role of the receiver first had to specify how much they would expect to receive in a dictator game. We used these first-order beliefs to incentivize the second-order belief elicitation from the dictators in the informed condition. As all receivers were paired with one dictator, they then received the amount the dictator allocated to them. Receivers in the informed condition were told that this money was allocated to them via a dictator game. As instructed to the dictators in the not-informed condition, receivers in the not-informed condition were not informed about where the money came from.

---

[4] In the informed condition, 75% of participants answered both questions correctly and the rest only one question correctly. In the not-informed condition, 88% of the participants an-

3.3 Participants

We recruited 299 participants on academic Prolific for the dictator role. A total of 281 participants took part at both measurement time points (of which 147 (52%) were in the not-informed condition). We excluded participants who completed only the first part. Participants received a £3 show-up fee plus a variable bonus between £0 and £5.5 for completing both parts of the study.

We pre-screened for participants who currently lived in the United States and had U.S. nationality. This was done to increase the comparability of the personal norm data with Krupka and Weber's (2013) injunctive norm data. Participants ranged in age from 19 to 76 years, with a mean age of 38 years (*SD* = 12.5). Forty-two participants (15%) were enrolled at a university at the time

swered both questions correctly, 9% only one question correctly, and 3% failed both questions.

of taking part in the study. The sample consisted of 115 females, 161 males, and 5 participants who did not fit into either of these categories or preferred not to say.

In addition to the participants in the dictator role, a total of 207 recipients (for the dictator games) were also recruited on academic Prolific. They received a £0.2 show-up fee and a mean bonus payment of £1.64 (depending on the decisions of the dictators).

4 Results

4.1 Motive Elicitation

We first report the results of the elicitations of the di↵erent motives. Fig. 1A plots the distribution of second-order beliefs about the recipients' expectations (what the dictators thought the recipients expected them to give) per condition. As expected, these distributions look very similar and there is no

significant mean di↵erence between conditions (M_{informed} = 3.5, $M_{\text{not-informed}}$ = 3.2), $t(278.6)$ = 0.82, p = .411. This was expected because also in the not-informed condition, participants were asked to guess the expectation of a recipient who was informed about the game. Since there is no significant di↵erence in second-order beliefs between the two conditions, any di↵erence in guilt aversion between conditions in our models should reflect di↵erences in the underlying mental processes, that is, letting one's behavior be influenced by these beliefs or not.

Fig. 1B shows the mean personal norm rating (personal appropriateness rating) per action per condition. The inverse U-shaped distribution is similar to the distribution of injunctive norm ratings in Krupka and Weber (2013) (see Fig. 1C for a comparison). As in Krupka and Weber (2013), the 50:50 split was perceived as most appropriate, and it was more appropriate to exit the game than to play the game and give nothing. A noteworthy di↵erence between our data and the data from Krupka and Weber (2013) is that the range of our data is smaller (i.e., especially the spread of the mean ratings between giving 0 points and giving 5 points is more pronounced in Krupka and Weber, 2013). This might be because we ran our study online. A study by K¨onig-Kersting (2021) found the same decreased range as compared to Krupka and Weber (2013). The reduced range in the norm ratings might be a reason why in our case, in contrast to Krupka and Weber (2013), norms do not do a good job in explaining behavior in the dictator game with an exit option (reported later). Looking at how the norm ratings di↵er between the two conditions, we see that there are no significant di↵erences except for the decision to give nothing to the recipient, which was seen as more appropriate than in the not-informed condition.

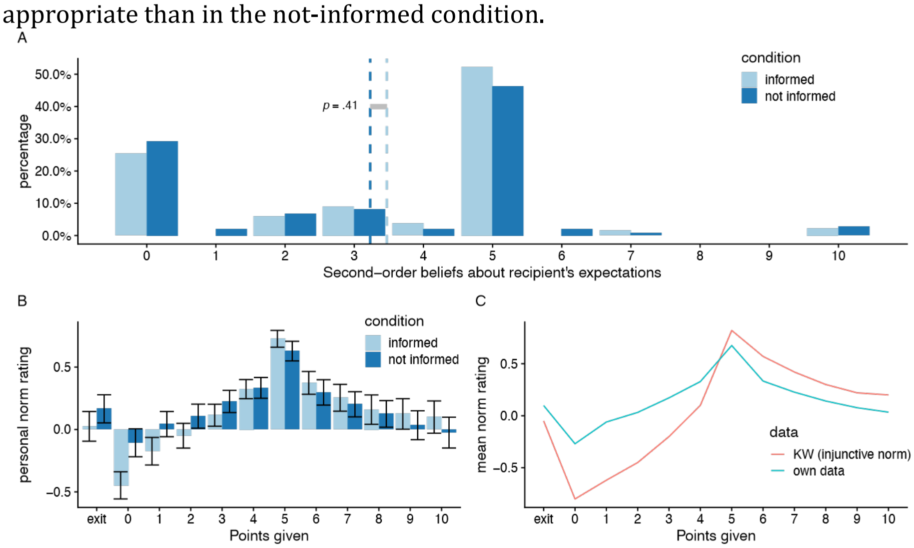


Figure 1: (A) Distribution of second-order beliefs. Dashed vertical lines represent the means, with the p value from a t test. (B) Mean personal norm ratings with 95% confidence intervals. (C) Mean norm rating comparison with data from Krupka and Weber (2013).

4.2 Dictator Game Without an Exit Option (T1)

In the dictator game without an exit option (T1), participants in the informed condition gave significantly more (M = 3.3 points) than participants in the not-informed condition (M = 2.0 points), $t(276.82) = 4.82$, $p < .001$. This is consistent with hypothesis 1, namely, that guilt aversion leads to more giving if no exit option is available. As seen in Fig. 2A, more participants chose the 50:50 split in the informed condition, and fewer participants chose to keep the whole endowment for themselves. In line with the results of Dana et al. (2006), the mean di↵erence in giving indicates that guilt aversion leads to more prosocial allocation decisions, since the only di↵erence between the two conditions was that the prerequisites for experiencing guilt (recipients were informed about the choice of the dictator) were met in the informed but not in the not-informed condition.

Table 2 shows the results of the estimated mixed-logit models for the behavior in the dictator game. All predictors are standardized to make comparisons between predictors meaningful. We fitted mixed-logit models because they allow for individual heterogeneity in preferences. When describing the results we focus on models that account for both motives and the payo↵ of an action (i.e., models 3 and 4 for the informed and models 7 and 8 for the not-informed condition). These are also the models that have the lowest Bayesian information criterion (BIC) values. The coe cients of the models represent the estimated

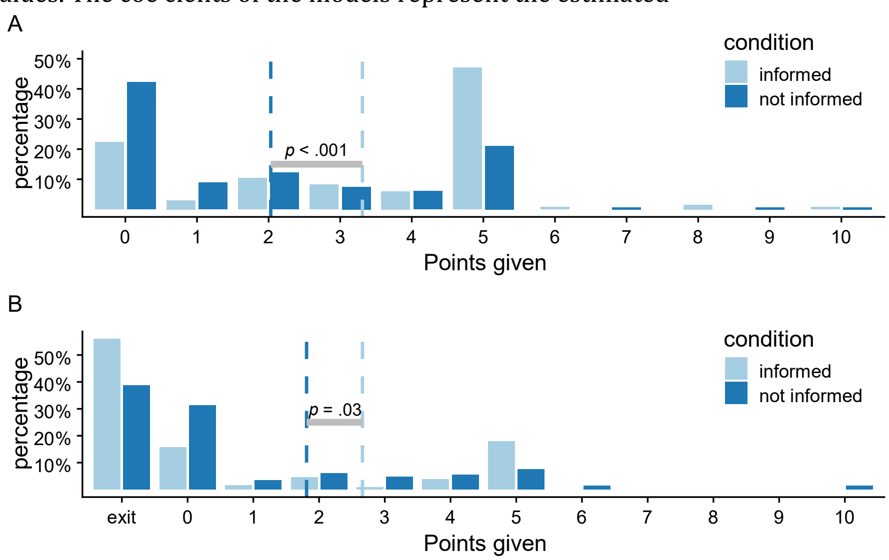


Figure 2: (A) Points given in the dictator game without an exit option (T1). (B) Points given in the dictator game with an exit option (T2). Vertical dashed bars represent the average points given by condition, with *p* values from *t* tests.

utility weights of the corresponding utility specification [i.e., equations (1), (2), and (3)]. The results reveal that the payo↵ of an alternative was the most important choice predictor since it has the highest utility weight. Guilt has a negative utility weight, which is in line with the guilt-aversion motive. Guilt had a larger negative e↵ect on utility in the informed than in the not-informed condition. Although we can by design rule out guilt in the not-informed condition (because recipients were not informed about the game), guilt still had some e↵ect on the choices in this condition. But the magnitude is considerably smaller, which is also evident in the fact that guilt was correlated with personal norms only in the informed condition. This is also reflected in the BIC: Only in the informed condition does the model with the guilt * self-image interaction have a lower BIC than the model without the interaction. The self-image coe cient is positive and significant in all models. Participants thus on average get positive utility from choosing an action that aligns with their personal norms. There is heterogeneity in preferences since in most models the variation in the random parameters is significant. Finally, the interaction between guilt and self-image is positive but only significant in the informed condition. It thus seems important to account for both motives if both motives can be present by design. Indeed, the significant negative e↵ect of guilt on the utility of an action vanishes in the not-informed condition once we account for the interaction of guilt and self-image.

Table 2: Mixed-Logit Models of Behavior in the Dictator Game

	Informed condition									Not-informed condition		
			(3)	(4)	(5)	(6)	(7)					
			3.869***	4.051***	2.439***	2.622***	3.540***					
			(0.585)	(0.620)	(0.508)	(0.385)	(0.579)					
			2.147***	2.437***	0.923*		0.752*					
			(0.522)	(0.675)	(0.397)		(0.382)					
			2.924***	3.390***		2.014***	2.035***					
			(0.532)	(0.649)		(0.369)	(0.371)					
Guilt * self-image								0.925⇤				0.127
								(0.417)				(0.209)
SD Payo↵	0.989⇤	0.446	0.001					0.041	0.785	0.884	0.203	0.196
	(0.461)	(0.693)	(0.780)					(0.833)	(0.517)	(0.453)	(1.796)	(1.699)
SD Guilt	2.059⇤⇤		2.133⇤⇤					3.296⇤⇤	1.875⇤⇤⇤		1.005⇤	1.045⇤
	(0.656)		(0.670)					(1.053)	(0.538)		(0.442)	(0.453)
SD Self-image		1.267⇤⇤	2.451⇤⇤⇤					3.012⇤⇤⇤		1.904⇤⇤⇤	1.987⇤⇤⇤	2.156⇤⇤
		(0.477)	(0.599)					(0.770)		(0.527)	(0.568)	(0.672)
Observations	134	134	134					134	147	147	147	147
Log Likelihood	268.464	228.291	203.949					199.952	262.349	216.915	210.310	210.110
Bayesian Inf. Crit.	556.519	476.174	437.284					434.189	544.659	453.791	450.563	455.153

Note: Note: The random coe cients follow a normal distribution. BIC = Bayesian information criterion. ⇤*p* < .05; ⇤⇤*p* < .01; ⇤⇤⇤*p* < .001

4.3 Dictator Game With an MPL Exit Option (T2)

We now turn to the dictator game with an MPL exit option. Fig. 3 plots the distribution of the exit reservation prices per condition. Participants in the informed condition were on average willing to exit the game for 6.4 points and were thus willing to forego 36% of the endowment they could potentially keep in the dictator game. Participants in the not-informed condition were on average willing to exit the game for 7.8 points and were thus willing to forego 22% of the potential endowment in the dictator game. As predicted by hypothesis 2, the mean exit reservation price was significantly lower in the informed condition, t(272.25) = 3.21, p = .001, meaning that participants in this condition were more willing to exit the game compared to participants in the not-informed condition. As a buying price was randomly drawn from a uniform distribution, more participants ended up with the exit option in the informed condition (75 participants, 56%) than in the not-informed condition (57 participants, 39%), ^{2}(1) = 7.64, p = .006. Given that the only di↵erence between the two conditions was that the prerequisites for experiencing guilt were met only in the informed condition, this di↵erence in exit reservation prices and resulting exit frequency indicates that people exit the dictator game to avoid feeling guilty (which is in line with Dana et al., 2006).

If participants indeed took the exit option to avoid guilt, the correlation between the points given in the dictator game at T1 and the reservation price of exiting the game at T2 should be moderated by the condition. This is because,

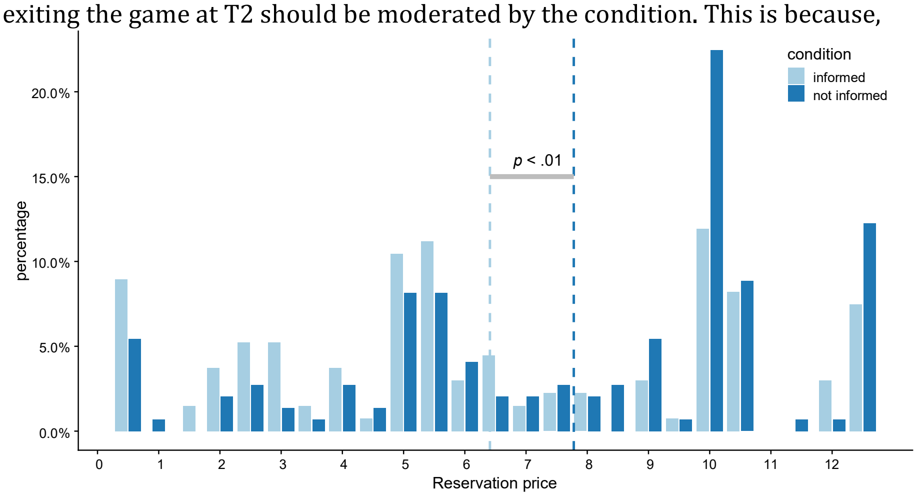


Figure 3: Exit option reservation price (in points) by condition.

as we proposed in hypothesis 1, guilt aversion likely drives giving if no exit option is available. But this should be the case only in the informed condition, since by design, guilt should have no influence in the not-informed condition. And indeed we find that the amount given in the dictator game at T1 was negatively correlated with the exit reservation price, but only for the informed condition (Table 3, model 1). Thus, the more participants in the informed condition gave in the dictator game at T1, the lower was their reservation price for exiting the game at T2 (i.e., the more money they were willing to forego for not having to play the game). This suggests that participants in the informed condition gave not because they wanted to but because to avoid feeling guilty, they had to give once they were in the sharing situation. This is consistent with Cain et al.'s (2014) account of giving in. Since we see this behavior only in the informed condition, guilt (aversion) is the likely driving force behind it.

Only 30.6% of participants in the informed and 44.9% in the not-informed condition had exit reservation prices that are consistent with conventional selfish or social preferences (i.e., exit reservation price dictator game endowment). This di↵erence is significant, $^{2}(1) = 5.49$, $p = .02$.

Fig. 2B shows the distribution of the chosen actions in the dictator game with an exit option. The distribution of the points given looks very similar to that in the dictator game without an exit option. Conditional on playing the game, participants in the informed condition on average gave significantly more (M = 2.7 points) to the recipient than participants in the not-informed condition (M = 1.8 points), $t(126.76) = 2.21$, $p = .03$. As in the dictator game without an exit option, more participants kept the whole endowment for themselves in the not-informed condition, whereas participants in the informed Table 3: Regression Analyses

	Dependent variable					
	Reservation price			Points given in DG with exit option (T2)		Points given in DG with exit option (T2) (exiting coded as 0)
	(1)	(2)	(3			
	0.113 (0.128)	0.762*** (0.069)	0.49 (0.0			
Informed condition	0.087 (0.663)			0.187 (0.377)		0.226 (0.3 48)
DG points given (T1) * informed condition	0.394* *** (0.187)			0.012 (0.109)		0.239⇤ (0.0 98)
Constant	8.002 (0.387)			0.313 (0.208)		0.111 (0.2 03)
Observations	281			149		281
R^2				0.594	0.191	
Adjusted R^2				0.586	0.183	
Residual Std. Error	3.459 (df = 277)			0.792*** (df = 3; 145)	21.838*** (df =	1.816 (df = 277)
F Statistic	8.559					

Note: DG = Dictator game. ⇤$p < .05$; ⇤⇤$p < .01$; ⇤⇤⇤$p < .001$

condition more often went for the 50:50 split.

Results of the mixed-logit models used to fit the data from the dictator game with an exit option are shown in Table 4. Again, all predictors were standardized to make comparisons between predictors meaningful. We again focus on models that account for both motives and the payo↵ of an action (models 3 and 4 for the informed and models 7 and 8 for the not-informed condition). These are also the models that have the lowest BIC values. In contrast to the standard dictator game (T1), guilt now has a larger utility weight than self-image. The (absolute) utility weight of guilt was even higher than that of the payo↵ of an action. This may seem strange at first, but it is in line with the fact that participants were willing to give up a significant amount of money to avoid guilt (i.e., to avoid playing the game). The utility weight of guilt is significant in all models, whereas the self-image predictor fails to reach significance in any model. These models show that for the dictator game with an exit option, avoiding guilt was even more important than the payo↵ of a given action, whereas self-image had an e↵ect only via the interaction with guilt. Regarding the importance of accounting for both motives simultaneously, we see that the interaction between self-image and guilt is now significantly positive in both conditions.

Participants for whom the game was implemented behaved very similarly to participants in the dictator game in the first session, $r(147) = 0.77$, $p < .001$. The strength of the correlation did not di↵er between conditions (see model 2 in Table 3). However, when we code the exit option as giving zero points, the correlation drops to $r(279) = .40$, $p < .001$. Using this coding, a multiple regression revealed that the correlation between the two games was significantly weaker in the informed condition (see model 3 in Table 3). In our view, these results have implications for the external validity of the dictator game, since outside of the lab, situations that allow for prosocial behavior can often be avoided. We therefore assume that the external validity of the dictator game with an exit option is higher than that of the dictator game without an exit option.

Although participants for whom the game was implemented behaved similarly to participants in the first game, the exit option left recipients significantly worse o↵, since they were left empty-handed when the dictator exited. Whereas in the standard dictator game, recipients in the informed condition received 3.3 points on average, in the dictator game with an exit option they received on average 2.14 points less, $t(133) = 9.77$, $p < .001$. This drop is less pronounced in the not-informed condition, where recipients on average received 2.03 points in the standard dictator game and 0.93 points less in the dictator game with an exit option, $t(146) = 5.48$, $p < .001$. So whereas recipients in the informed condition were significantly better o↵ when there was no exit option, this was no longer the case once the exit option was present, $t(277.23) = 0.26$, $p = .79$. This suggests that guilt has no e↵ect on aggregate giving when avoiding the sharing situation is possible: Guilt leads people to give more if they enter the game (intensive margin), but this is o↵set by more people avoiding the game (extensive margin). In consequence, guilt had an overall negative e↵ect on aggregate welfare, since participants in the informed condition were leaving significantly more money on the table (lower exit reservation price) than participants in the not-informed condition.

Table 4: Mixed-Logit Models of Behavior in the Dictator Game With an Exit Option

	Informed condition									Not-informed condition		
				(4)	(5)	(6)	(7)	(8)				
				0.581***	0.953***	0.639***	0.981***	0.999*				
				(0.166)	(0.171)	(0.100)	(0.157)	(0.17				
				2.787*	1.231*		1.289*	1.20				
				(1.132)	(0.496)		(0.525)	(0.55				
Self-image		0.318	0.210	0.548						0.169	0.071	0.278
		(0.194)	(0.309)	(0.373)						(0.138)	(0.201)	(0.234)
Guilt * Self-image				0.963⇤								0.662⇤
				(0.425)								(0.282)
SD Payo↵	0.002	0.172	0.028	0.124					0.403	0.009	0.081	0.081
	(0.201)	(0.782)	(0.953)	(1.278)					(0.321)	(0.478)	(0.880)	(1.201)
SD Guilt	2.962⇤⇤		3.787⇤⇤⇤	3.792⇤⇤					2.851⇤⇤⇤		3.120⇤⇤⇤	3.357⇤⇤⇤
	(1.003)		(1.150)	(1.346)					(0.829)		(0.880)	(0.996)
SD Self-image		1.369⇤⇤	2.501⇤⇤⇤	2.665⇤⇤⇤						0.725⇤	1.309⇤⇤⇤	1.420⇤⇤⇤
		(0.454)	(0.594)	(0.732)						(0.315)	(0.380)	(0.397)
Observations	134	134	134	134					147	147	147	147
Log Likelihood	311.762	321.635	295.824	292.004					322.263	337.419	316.441	312.565
Bayesian Inf. Crit.	643.116	662.861	621.036	618.293					664.487	694.799	662.824	660.063

Note: The random coe cients follow a normal distribution. BIC = Bayesian information criterion. ⇤*p* < .05; ⇤⇤*p* < .01; ⇤⇤⇤*p* < .001

The models in Table 4 tell us how well the motives can explain behavior in the dictator game with an exit option as a whole, not just with regard to the decision of whether to exit the game. Because the exit reservation price influences the probability that a participant has to play the game, we can test whether these motives influence the decision to exit the game by letting them interact with the exit reservation price. If guilt (aversion) and self-image concerns lead participants to exit the dictator game, the e↵ect of these motives should interact with the exit reservation price. For the guilt motive we thus expect that the more guilt averse a person is, the lower the exit reservation price should be (since a low exit reservation price increases the probability of exiting; hypothesis 2). For the self-image motive, we expect that the more important a positive self-image is to a person, the lower the exit reservation price should be (hypothesis 3). We incorporate the exit reservation price into the model by assuming that the variation in coe cients of guilt and personal norm across individuals is influenced by the exit reservation price of an individual[5] (see Table

[5] For example, for guilt $_{guilt,i} = {}_1 +$ ⇤ *reservationprice* + ℋwhere ℋ⇠ $N(0,1)$; see Sarrias and Daziano (2017) section 3.4.

5).

Table 5: Mixed-Logit Models of Behavior in the Dictator Game With Exit Option Reservation Price

	Informed co ndition				Not-informed condition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Payoff	0.571***	0.179	0.701***	0.688***	1.097***	0.638***	1.242***	1.250***
	(0.116)	(0.117)	(0.187)	(0.187)	(0.204)	(0.100)	(0.255)	(0.264)
Guilt	6.070**		7.422***	7.230**	4.933***		5.204***	5.175***
	(1.996)		(2.210)	(2.280)	(1.338)		(1.324)	(1.398)
Guilt.reservation.price	0.523**		0.689**	0.684**	0.504***		0.549***	0.554***
	(0.183)		(0.210)	(0.221)	(0.139)		(0.140)	(0.150)
Self-image		0.226	1.324⇤	0.984		0.015	0.883	0.653
		(0.385)	(0.615)	(0.659)		(0.339)	(0.535)	(0.563)
Self-image.reservation price		0.088	0.247⇤⇤	0.227⇤		0.023	0.131⇤	0.113
		(0.057)	(0.091)	(0.095)		(0.039)	(0.065)	(0.068)
Guilt * Self-image				0.622				0.445⇤
				(0.362)				(0.221)
SD Payoff								
SD			2.699**	2.785**	1.617**		1.424*	1.509*
SD Self-image		1.381**	2.235***	2.377***		0.722*	1.456***	1.544***
		(0.451)	(0.579)	(0.638)		(0.318)	(0.408)	(0.434)
Observations	134	134	134	134	147	147	147	147
Log Likelihood	300.278	320.246	278.675	276.801	304.981	337.239	295.477	292.851
Bayesian Inf. Crit.	625.045	664.980	596.533	597.682	634.914	699.430	630.877	630.617
	0.002	0.355	0.304	0.308	0.568	0.010	0.547	0.556
	(0.213)	(0.464)	(0.641)	(0.638)	(0.317)	(0.475)	(0.431)	(0.452)
Guilt	2.392⇤⇤							
	(0.825)		(0.893)	(0.954)	(0.586)		(0.592)	(0.670)

Note: The random coe cients follow a normal distribution. BIC = Bayesian information criterion. Guilt.reservation price and Self-image.reservation price represent the e↵ect of the reservation price on the mean of these random parameters. ⇤$p < .05$; ⇤⇤$p < .01$; ⇤⇤⇤$p < .001$ Again, all predictors are standardized to make comparisons between predic-

tors meaningful and we focus on the models where both motives are present (these are the models with the lowest BIC values). Incorporating the exit reservation price starkly increases the (absolute) utility weight of guilt. The exit reservation price can explain some of the heterogeneity of the guilt predictor, which is evident in the decreased standard deviation of that predictor and the significant interaction between guilt and the exit reservation price. This positive interaction indicates that the more guilt averse a participant was, the lower was their exit reservation price, which confirms hypothesis 3. The utility weights of the self-image predictor turn negative, but only significantly so for model (3). Self-image interacts with the exit reservation price, but the interaction was not as strong as it was for guilt and the exit reservation price. Consistent with this, the drop in heterogeneity in self-image is only subtle. In contrast to hypothesis 4, the interaction between the exit reservation price and self-image was positive. Participants who cared about their self-image were thus more likely to play the game than to take the exit option (because they on average had a higher exit reservation price). Finally, we see that the interaction between self-image and guilt is positive and significant in the not-informed condition. Models that account for both motives also fit the data best according to the BIC.

The results on guilt (aversion) are quite consistent: Participants were less likely to take actions that (would) make them feel guilty. In the dictator game, this led to more giving, but also to greater avoidance of the game when avoidance was possible. Results for the self-image motive are less consistent. When no exit option was present (i.e., dictator game at T1), participants took actions that led to a positive self-image. But this behavior was no longer observed once the exit option was available. This suggests that participants who took the exit option cared less about their self-image than participants who played the game. To test this we fitted the models in Table 4 only for those who did not take the exit option, and indeed the self-image predictor now had a positive and significant utility weight (see Table A1 in the appendix). Thus, instead of being concerned about their self-image, people who took the exit option seem to have cared less about their self-image than people who played the game.

Table 6: Guilt and Shame Sensitivity

	Informed condition		Not-informed condition	
	Game			
	DG	DG exit	DG	DG exit
	(1)	(2)	(3)	(4)
Payo↲	4.065⇤⇤⇤	0.581⇤⇤⇤	3.532⇤⇤⇤	1.009⇤⇤⇤
	(0.622)	(0.165)	(0.528)	(0.193)
Guilt	0.290	5.260	0.378	2.927
	(4.650)	(5.131)	(1.903)	(3.470)
Self-image	5.195⇤	1.181	0.366	0.774
	(2.118)	(1.494)	(1.516)	(1.147)
Guilt * Self-image	0.937⇤	0.880⇤	0.133	0.697⇤

	(0.426)	(0.414)	(0.208)	(0.288)
Guilt.guilt score	0.513	1.892	0.078	0.417
	(1.090)	(1.278)	(0.436)	(0.806)
Self-image.shame score	0.543	0.203	0.732	0.320
	(0.572)	(0.428)	(0.470)	(0.339)
SD Payo↵	0.024	0.138	0.030	0.165
	(0.884)	(1.115)	(1.551)	(0.899)
SD Guilt	3.433⇤⇤	3.797⇤⇤	1.111⇤	3.243⇤⇤⇤
	(1.170)	(1.289)	(0.454)	(0.969)
SD Self-image	2.953⇤⇤⇤	2.610⇤⇤⇤	2.065⇤⇤	1.455⇤⇤⇤
	(0.757)	(0.701)	(0.670)	(0.405)
Observations	134	134	147	147
Log Likelihood	199.453	290.546	208.748	311.981
Bayesian Inf. Crit.	442.986	625.173	462.411	668.877

Note: DG = dictator game; DG exit = DG with exit option; BIC = Bayesian information criterion. The random coe cients follow a normal distribution. Guilt.guilt score and Self-image.shame score represent the e↵ect of the guilt sensitivity and shame sensitivity on the mean of these random parameters. ⇤*p* < .05; ⇤⇤*p* < .01; ⇤⇤⇤*p* < .001

The results of the models where we included guilt and shame proneness as measured with the TOSCA-3 are shown in Table 6. The guilt and shame subscales have Cronbach's alphas of 0.76 and 0.78, respectively. When incorporating the guilt and shame score into the logit models, neither guilt nor shame sensitivity significantly a↵ected choice behavior. In contrast to Bellemare et al. (2019), we found no correlation between the economic (second-order beliefs) and psychological (TOSCA-3 guilt score) ways of measuring guilt, $r(279) = 0.04$, $p = .52$.

5 Discussion and Conclusion

Past studies showed that both guilt aversion Dana et al. (2006) and self-image concerns (Grossman and van der Weele, 2017; Tonin and Vlassopoulos, 2013) can explain why participants choose the exit option in a dictator game. However, no study so far elicited both motives and pitched the explaining power of the two motives against each other. This is problematic, as the two motives could be correlated, which could potentially lead to an overestimation of any given motive. Using the dictator game and a between-subjects design where by design, guilt should occur in only one condition, we therefore set out to test whether guilt aversion or self-image concerns can better explain why some people avoid sharing situations. Our results show that in the dictator game with an exit option, guilt has a larger effect on an action's utility than self-image. Both predictors are moderated by the exit reservation price. As predicted, the interaction is positive for the guilt motive. Thus, the more guilt averse participants were, the more money they were willing to give up to avoid the game (lower exit reservation price). This is consistent with the reasoning that people take the exit option to avoid guilt. In contrast to our hypothesis, the interaction between the utility weight of a positive self-image and the exit reservation price is also positive. Thus, the less participants wanted to avoid the game (higher exit reservation price), the more utility they got from a positive self-image. The results concerning the self-image motive in the dictator game with an exit option offer an interesting contrast to previous studies that used injunctive norms to explain behavior in this game. In contrast to Krupka and Weber (2013) who used injunctive norms to explain behavior in the dictator game with an exit option, personal norms cannot explain behavior in this game in our study.

Coming back to the question of whether people share because they think it is the right thing to do (personal norm/self-image) or because to avoid feeling guilty they must do so (guilt aversion), our results suggest that a large share of people only give because guilt aversion compels them to do so. People seemed to be willing to give up a lot of money to avoid the sharing situation (dictator game in our case) and this willingness to give up money was correlated with how guilt averse they were in the dictator game. Regarding the importance of accounting for both motives simultaneously, we saw that this is indeed important, as (1) models that accounted for both motives fit the data best (lowest BIC values) and (2) the two motives were often correlated.

Given that guilt aversion leads people to give up money to avoid the dictator game (via a low exit reservation price), and given that guilt does not lead to more aggregate sharing once we account for this increased avoidance, we conclude that guilt as a means to boost prosocial behavior should be used with caution. It should be used only if the guilt-inducing situation cannot be avoided or the costs of avoidance are very low. Otherwise, the overall effect of guilt on welfare is likely negative because avoidance in this case can lead to welfare-decreasing activities. This also has practical implications. For instance, charities should account for the fact that using guilt to attract donations (e.g., Basil et al., 2006) could potentially backfire owing to increased avoidance. Past research has already shown that people do not like being asked for donations (Andreoni et al., 2017; Adena and Huck, 2019). Our results suggest that this effect should be even stronger when fundraisers deliberately use guilt to catalyze giving. Moreover, recent research in organizational psychology has proposed encouraging guilt responses after employees' failure to generate benefits for employees and organizations (e.g., Bohns and Flynn, 2013). According to our results, this strategy may suffer from the possibility that guilt drives employees to also avoid situations in which prosocial behavior is possible. In line with this, future research could examine whether employees also engage in harmful behavior to avoid potentially guilt-inducing situations.

The differential impact of guilt (aversion) on the extensive and intensive margins also has important implications for the external validity of social preference games. Since avoidance of these games is often not possible in the lab, guilt leads to more giving. However, in the real world, avoidance is (often) possible, and the people who gave because of guilt (aversion) in the lab will likely avoid situations that allow for sharing to avoid guilt outside of the lab. Social preference games that do not allow for sorting thus overestimate prosociality. This might explain why some studies (e.g., Galizzi and Navarro-Martinez, 2019) find poor external validity for social preference games. The results of Cappelen et al. (2017) support this claim: Only in the condition where by design guilt was not possible did the behavior in the game correlate significantly with charitable giving outside of the lab. Future research could test whether offering participants an exit option in the lab improves the external validity of social preference measures such as the dictator game.

It is interesting to see that personal norms cannot explain why people take the exit option, but injunctive norms can (Krupka and Weber, 2013). This is surprising because the enforcement of injunctive norms, but not personal norms, depends on other people who observe norm violation (Anderson and Dunning, 2014). Maybe people have internalized the injunctive norms, such that enforcement no longer depends on others observing the behavior (Gross and Vostroknutov, 2022). The different explanatory power of personal and injunctive norms on behavior in the dictator game with an exit option is congruent with the results of Bašić and Verrina (2020) who showed that personal norms and injunctive norms measure different things and do not have to align on an individual basis. This suggests that people who take the exit option think that most others think this is an appropriate thing to do (results of Krupka and Weber, 2013), but they themselves think it is rather inappropriate (our results). Future research could test this by following Bašić and Verrina (2020) and eliciting both personal and injunctive norms.

Our study suffers from similar limitations to those of previous studies that used a dictator game with an exit option. The main limitation is that one cannot rule out that participants exited the game for motives other than guilt aversion and self-image concerns. As summarized nicely by Exley and Kessler (2021), it could be that participants (1) may have had a preference to avoid interpersonal trade-offs, (2) were so inattentive that they randomly chose options, and/or (3) were so confused that they did not understand the value of taking the exit option or not. We tried to circumvent the second and third possibilities by highlighting to participants that we would ask comprehension questions and by providing feedback on those questions, but we cannot rule out that participants preferred the exit option because it allowed them to avoid interpersonal trade-offs.

References

Adena, M. and Huck, S. (2019). Online fundraising, self-image, and the longterm impact of ask avoidance. *Management Science*.

Amelio, A. and Zimmermann, F. (2023). Motivated memory in economicsmdash;a review. *Games*, 14(1).

Andersen, S., Harrison, G. W., Lau, M. I., and Rutstr¨om, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Anderson, J. E. and Dunning, D. (2014). Behavioral norms: Variants and their identification. *Social and Personality Psychology Compass*, 8(12):721–738.

Andreoni, J., Rao, J. M., and Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, 125(3):625–653.

Balafoutas, L. and Fornwagner, H. (2017). The limits of guilt. *Journal of the Economic Science Association*, 3(2):137–148.

Balafoutas, L. and Sutter, M. (2017). On the nature of guilt aversion: Insights from a new methodology in the dictator game. *Journal of Behavioral and Experimental Finance*, 13:9–15.

Baˇsi´c, Z. and Verrina, E. (2020). Personal norms — and not only social norms — shape economic behavior. *Discussion Papers of the Max Planck Institute for Research on Collective Goods 2020/25*.

Basil, D. Z., Ridgway, N. M., and Basil, M. D. (2006). Guilt appeals: The mediating e↵ect of responsibility. *Psychology & Marketing*, 23(12):1035–1054.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97(2):170–176.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Battigalli, P. and Dufwenberg, M. (2020). Belief-dependent motivations and psychological game theory. CESifo Working Paper 8285, Munich.

Bellemare, C., Sebald, A., and Suetens, S. (2019). Guilt aversion in economics and psychology. *Journal of Economic Psychology*, 73:52–59.

Bohns, V. K. and Flynn, F. J. (2013). Guilt by design: Structuring organizations to elicit guilt as an a↵ective reaction to failure. *Organization Science*, 24(4):1157–1173.

Bracht, J. and Regner, T. (2013). Moral emotions and partnership. *Journal of Economic Psychology*, 39:313–326.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, 94(1):32–37.

Burks, S. V. and Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1):203–217.

Cain, D. M., Dana, J., and Newman, G. E. (2014). Giving versus giving in. *The Academy of Management Annals*, 8(1):505–533.

Cappelen, A. W., Halvorsen, T., Sørensen, E. Ø., and Tungodden, B. (2017). Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Association*, 15(3):540–557.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015–1026.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

Daniels, M. A. and Robinson, S. L. (2019). The shame of it all: A review of shame in organizational life. *Journal of Management*, 45(6):2448–2473.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1):1–56.

Edwards, J. T. and List, J. A. (2014). Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. *Journal of Public Economics*, 114:1–13.

Elster, J. (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge University Press, Cambridge.

Exley, C. L. and Kessler, J. B. (2021). Information avoidance and image concerns. Working Paper 28376, National Bureau of Economic Research.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature human behaviour*, 2(7):458–468.

Galizzi, M. M. and Navarro-Martinez, D. (2019). On the external validity of social preference games: A systematic lab-field study. *Management Science*, 65(3):976–1002.

G¨odker, K., Jiao, P., and Smeets, P. (2021). Investor memory. *Available at SSRN 3348315*.

Gross, J. and Vostroknutov, A. (2022). Why do people follow social norms? *Current Opinion in Psychology*, 44:1–6.

Grossman, Z. and van der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173– 217.

Hauge, K. E. (2016). Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology*, 54:35–43.

Knutsson, M., Martinsson, P., and Wollbrant, C. (2013). Do people avoid opportunities to donate?: A natural field experiment on recycling and charitable giving. *Journal of Economic Behavior Organization*, 93:71–77.

K¨onig-Kersting, C. (2021). On the robustness of social norm elicitation. *Working Papers in Economics and Statistics*.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

McFadden, D. (1973). Conditional logit analysis ofqualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Sarrias, M. and Daziano, R. (2017). Multinomial logit models with continuous and discrete individual heterogeneity in r: the gmnl package. *Journal of Statistical Software*, 79(2):1–46.

Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in experimental social psychology*, volume 10, pages 221–279. Elsevier.

Strohminger, N. and Nichols, S. (2014). The essential moral self. *Cognition*, 131(1):159–171.

Tangney, J. P., Dearing, R., Wagner, P. E., and Gramzow, R. (2000). *The Test of Self-Conscious A↵ect3 (TOSCA-3)*. George Mason University, Fairfax, VA.

Tangney, J. P., Stuewig, J., and Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual review of psychology*, 58:345–372.

Tangney, J. P., Wagner, P., and Gramzow, R. (1992). Proneness to shame, proneness to guilt, and psychopathology. *Journal of Abnormal Psychology*, 101(3):469–478.

Tonin, M. and Vlassopoulos, M. (2013). Experimental evidence of self-image concerns as motivation for giving. *Journal of Economic Behavior & Organization*, 90:19–27.

Trachtman, H., Steinkruger, A., Wood, M., Wooster, A., Andreoni, J., Murphy, J. J., and Rao, J. M. (2015). Fair weather avoidance: unpacking the costs and benefits of "avoiding the ask". *Journal of the Economic Science Association*, 1(1):8–14.

Compassion for all: Real-world online donations contradict compassion fade

Dominik S. Meier
Center for Philanthropy Studies (CEPS)
University of Basel
Switzerland

Abstract

Research has shown that people are more likely to donate money to help a single victim rather than a group of victims. However, recent studies have been able to reverse this compassion fade e ect by presenting people with multiple donation appeals with di erent victim group sizes (joint evaluation) instead of just one donation appeal (separate evaluation). The reversal of this e ect when people evaluate multiple donation requests at once has important implications for fundraising. This study tests whether this e ect can be replicated in the field by using data from GoFundMe, the world's largest crowdfunding platform. When browsing projects on GoFundMe, people see multiple projects displayed at once, placing them in a joint evaluation context. Using the project campaign category and description to control for confounding, I find that there is indeed a positive e ect of the perceived victim group size on the amount of funds raised by a project.

Compassion for all: Real-world online donations contradict compassion fade Introduction

Online fundraising has become increasingly popular, as it provides access to a huge donor pool at very low costs (Hart, 2002). The promises of online fundraising have been successfully exploited by crowdfunding sites such as GoFundMe.com that allow people to set up crowdfunding campaigns on their online platform with only a few clicks. The $15 billion that has been collectively raised since 2010 on GoFundMe (GoFundMe, 2022), which is the largest donation-based crowdfunding website, speaks for the success of these platforms. Although the amount of funds raised on GoFundMe is impressive, most campaigns fall short of their fundraising targets (Kenworthy & Igra, 2022). This low success rate might be explained by the fact that traditional strategies intended to boost donations (e.g., Ruehle et al., 2021) might not work or even backfire on such platforms. This is because in contrast to more traditional means of fundraising, such as mail solicitation where potential donors often receive only one donation request at a time, potential donors on crowdfunding platforms can choose from a large number of projects to donate to. As noted by Erlandsson (2021), whether people evaluate one option separately or multiple options jointly has been very influential for research on judgment and decision making. For example, Erlandsson (2021) quotes evidence that shows that emotional reactions are more predictive of attitudes toward policies in separate evaluations (Ritov & Baron, 2011), while e ciency-related attributes are more predictive in joint evaluations (Bazerman et al., 2011; Caviola et al., 2014). This evidence led Erlandsson (2021) to test seven helping e ects (i.e., strategies that fundraisers can use to boost donations) both when people evaluated multiple donation requests at once (joint evaluation) and when they only evaluated one donation request (separate evaluation).

Erlandsson (2021) found that potential donors indeed prefer projects with di erent attributes depending on whether they only evaluate one project or multiple projects at once. For example, while research using separate evaluation found that donors prefer projects with a single identified victim to projects with multiple unidentified victims (Lee & Feeley, 2016), Erlandsson (2021) was able to reverse this e ect in the joint evaluation condition. This outcome is in line with another recent experimental study by Garinther et al. (2022). In contrast to previous studies that used a separate evaluation design, Garinther et al. (2022) found that people donated more to donation requests depicting larger victim groups than to donation requests depicting smaller victim groups when participants evaluated multiple donation requests at once. According to Garinther et al. (2022), it is the comparison of multiple donation requests with di erent depicted group sizes that leads to the positive e ect of depicted group size on giving. This result has important consequences for fundraising, since studies have traditionally concluded that larger victim groups attract smaller donations (see Butts et al. (2019) for a meta-analysis).

This study examines whether fundraisers that use crowdfunding can leverage the results of Erlandsson (2021) and Garinther et al. (2022) by manipulating the *perceived* victim group size. In lab studies that test the e ect of victim group size on giving, the perceived victim group size (i.e., how many people are depicted on the picture) usually corresponds with the real victim group size (i.e., the size of the group that will receive the donations) e.g., Garinther et al. (2022). However, in real-life donation requests, there is usually no direct correspondence between the size of the depicted victim group (e.g., a poor child from Sudan) and the size of the group that benefits from the donation (poor Sudanese children, in this example).[1] Thus, in this work, I attempt to test whether fundraisers can raise more funds by manipulating the size of the perceived victim group in a joint evaluation context (i.e., crowdfunding).

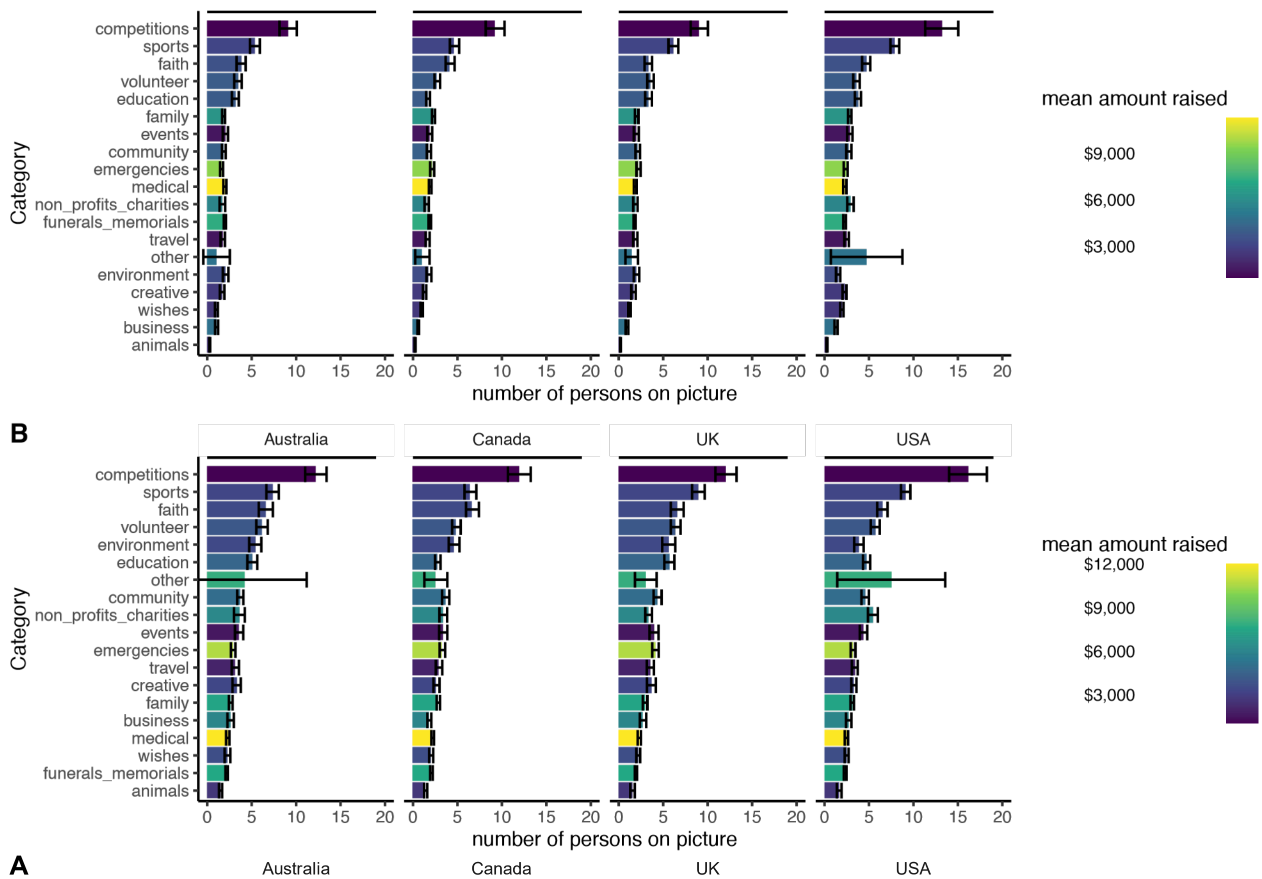
To test this hypothesis, I estimate the e ect of the depicted victim group size on giving in a real-world setting where people usually see multiple donation requests at once. I use data from more than 60,000 crowdfunding projects from GoFundMe, the world's largest social fundraising platform. When browsing fundraising projects on GoFundMe.com, people see multiple fundraising projects displayed in a grid (see Figure A1 in the appendix), which places them into a joint evaluation framework. According to Erlandsson (2021) and Garinther et al. (2022), we should thus observe a positive e ect of the number of persons depicted on a project's project profile picture on the funds acquired by the project.

Given the observational setting, I first need to identify the e ect of the perceived victim group size on the funds raised. To identify this e ect, I need to account for all confounders between the number of people depicted on a project profile picture and the amount of funds raised. The topic of a fundraising project is such a confounder. Whether the funds are raised for a sick child or a college football team likely influences both how many people are depicted on the profile picture and how much people will donate to the project. Fortunately, fundraising projects on GoFundMe must be assigned to one of 18 predefined categories (e.g., "medical", "sports"). Indeed, the category of a project correlates with both the amount of funds raised and the number of persons depicted on a project profile picture (see Figure 1).

To assess how robustly the category of a project controls for confounding, I also use the campaign description text to additionally control for the topic of a project. The campaign description is free text provided on the project's profile page that fundraisers can use to describe their project. Campaign descriptions have been shown to influence the success of crowdfunding projects (Kuo et al., 2022). I use document embeddings (Le & Mikolov, 2014; Reimers & Gurevych, 2019) and topic models (Blei, 2012) to encode this text into numbers that can then be used as controls.

Controlling for the category and the campaign description allows me to identify the e ect of the perceived victim group size on donations for projects that belong to the same category and have similar campaign descriptions. This places us close to an experimental design where we could vary the number of people depicted on the profile project picture while keeping the description of the project constant.

I use regression models and double machine learning (Chernozhukov et al., 2017) as a robustness check to estimate the e ect of the number of people on the profile picture on



B
A
Australia
Canada
UK
USA
Category
number of persons on picture
mean amount raised
$12,000
$9,000
$6,000
$3,000
competitions
sports
faith
volunteer
education
family
events
community
emergencies
medical
non_profits_charities
funerals_memorials
travel
other
environment
creative
wishes
business
animals
0 5 10 15 20

Figure 1

Mean number of persons on the project profile picture by project category. Error bars represent 95% confidence intervals. Figure A) with all projects, Figure B) only with projects that have at least one person on the project profile picture.

the amount of funds raised. Double machine learning (Chernozhukov et al., 2017) uses o -the-shelf machine learning algorithms to estimate causal e ects in the presence of potentially high-dimensional confounders. Double machine learning allows us to control for confounders (e.g., document embeddings) without making strong assumptions about the functional form of our model.

In contrast to the majority of the extant research that has mostly found a negative e ect on the victim group size on funds raised (Butts et al., 2019), I find no such negative e ect, or even a positive e ect, of the perceived victim group size on the amount of funds raised by crowdfunding projects. In contrast to past findings and in line with recent evidence from laboratory studies, it thus seems beneficial to increase the perceived victim group size in settings where potential donors evaluate multiple fundraisers at once. **Literature review**

E ect of victim group size on donations

There are two well-known e ects of victim group size on donations, namely, the identifiable victim e ect (Jenni & Loewenstein, 1997) and the compassion fade e ect (Västfjäll et al., 2014). For both e ects, the e ect of the victim group size on the funds raised is negative. The identifiable victim e ect refers to the tendency of individuals to provide more help to specific, identifiable victims than to anonymous (statistical) victims (Jenni & Loewenstein, 1997). For example, Kogut and Ritov (2005a) found that when asked to help sick children who need costly life-saving treatment, participants were more willing to contribute to a single child identified by age, name, and picture than to a single unidentified child or a group of unidentified children. While this e ect results in people donating less to larger victim groups, it mainly operates, as the name implies, through the identification of the victims (Lee & Feeley, 2016). Indeed, a meta-analysis by Lee and Feeley (2016) found that this e ect only works for a single victim and not for a group of victims. This finding is in line with the results of Kogut and Ritov (2005a) and Kogut and Ritov (2005b), who found that people donated more to a single identified victim than to a nonidentified victim, while there was no significant di erence between donations made to a group of identified victims and those made to a group of nonidentified victims.

In contrast to the identifiable victim e ect, the compassion fade e ect specifies a negative e ect of the depicted victim group size on willingness to donate that is directly caused by the size of the victim group. As mentioned by Butts et al. (2019), the compassion fade e ect has also been referred to as compassion fatigue (Figley, 1995), compassion collapse (Cameron, 2017), and psychic numbing (Slovic, 2007). Butts et al. (2019) also highlighted that it is important to note that compassion here refers to compassionate behavior (e.g., donating) and not to the emotion of feeling compassion. Butts et al. (2019) analyzed 41 studies in a meta-analysis and found that victim group size negatively a ected both helping intent and behavior (e.g., donations). They also found that anticipated positive a ect and perceived impact, which were negatively associated with victim group size, mediated this e ect.

One prominent explanation of the compassion fade e ect is that it is caused by numeracy limitations and biases in the basic a ective processing underlying the decision to help (Hamilton & Sherman, 1996; Slovic, 2007). Butts et al. (2019) called this the a ective bias perspective. This explanation is related to the identifiable victim e ect; it postulates that a single victim is depicted in more detail (i.e., with more information) than are groups, which elicits stronger emotional reactions (Hamilton & Sherman, 1996). In contrast to a single victim, a group of victims constrains people’s capacity for attention and imagery, which results in a fragmented representation of the victims and thus a weaker a ective response (Dickert & Slovic, 2009; Kogut & Ritov, 2005a).

The other prominent explanation postulates that people expect the needs of large groups to be potentially overwhelming and therefore engage in emotion regulation to prevent themselves from experiencing these overwhelming emotions (Cameron & Payne, 2011). According to this explanation, people regulate their emotions to maximize their personal goals while minimizing potential costs that may seem overwhelming (Butts et al., 2019). Butts et al. (2019) called this the motivated choice perspective and noted that this explanation aligns with the cost-reward model of helping (Dovidio et al., 1991) and past work on empathy avoidance (Shaw et al., 1994). The abovementioned work demonstrated that potential donors regulate their emotions to avoid feelings that will compel them to help when helping is foreseen as being too costly.

To explain the reversion of the identifiable victim effect in joint evaluation, Erlandsson (2021) referred to X. Li and Hsee (2019), who posited that attributes in decision situations can differ in both justifiability (whether people think the attribute should affect decisions) and in evaluability (how easily the attribute in itself can be understood). Erlandsson (2021) noted that the size of the victim group is a prime example of an attribute with a high level of justifiability (most people would agree that helping more people is better than helping fewer) but a low level of evaluability (without any comparison, it is difficult to judge whether four victims are few or many). However, evaluability is higher in joint evaluations than in separate evaluations (Erlandsson, 2021; Hsee, 1996). Hsee (1996) noted that there might be a greater level of uncertainty in judging the value of a hard-to-evaluate attribute (e.g., victim group size) in separate evaluations than in joint evaluations. Therefore, these factors could have less impact in separate evaluations than in joint evaluations (Hsee, 1996). In line with this, Hsee, Zhang, Wang, et al. (2013) showed that willingness to donate when one could save 200 rather than 100 polar bears was twice as high in joint evaluation, while there was no difference in separate evaluation.

While the evidence for the compassion fade effect is substantial (Butts et al., 2019), this evidence rests on some limitations. Mainly, as mentioned by Garinther et al. (2022), only a few studies have used designs that required participants to jointly evaluate donation requests with different victim group sizes. The meta-analysis from Butts et al. (2019) explicitly excluded such studies. The authors acknowledged this shortcoming by stating that designs with "separate evaluations do not adequately reflect the realistic settings in which people make donation decisions" (p. 27). This leads to the second limitation, namely, to the best of my knowledge, this e ect has never been tested using real-world donation data.

Regarding the first limitation, the few studies that have used joint evaluations to study the compassion fade e ect have found inconsistent results (Butts et al., 2019). As mentioned by Garinther et al. (2022), only one of the studies replicated the compassion fade e ect (Västfjäll et al., 2014, study 2). The other studies either found comparable donations to di erent victim group sizes (Kogut & Ritov, 2005b) or were even able to reverse the compassion fade e ect (Hsee, Zhang, Lu, et al., 2013; Kogut & Ritov, 2005b).

Garinther et al. (2022) took the existing inconsistencies in the design and results of these studies as motivation to systematically study the compassion fade e ect in joint evaluation conditions. Over multiple studies, the authors showed that when subjects saw multiple donation requests at once, either simultaneously or sequentially, they donated more to larger victim groups.

This finding has important practical implications since it contradicts the compassion fade e ect literature in a setting that "better mirror[s] real charitable giving contexts" (Butts et al., 2019, p. 27). Thus, in contrast to the majority of the extant literature on the relationship between victim group size and donations (Butts et al., 2019), fundraisers whose solicitations are evaluated jointly with other solicitations can potentially attract more donations by increasing the (perceived) victim group size. If this recommendation is externally valid, we should observe the following:

Hypothesis 1: There is a positive e ect of perceived victim group size (i.e., number of persons on the project profile picture) on the amount of funds raised by the project.

Success Factors of Donation based Crowdfunding Campaigns

To understand what drives a crowdfunding campaign's success we need to understand what motivates people to give to those campaigns and how crowdfunding campaigns can tap into these motivations. This short review largely draws on two excellent reviews by van Teunenbroek and Dalla Chiesa (2022) and van Teunenbroek et al. (2022). van Teunenbroek and Dalla Chiesa (2022) summarise motives that lead people to donate to crowdufunding campaigns. Many of the mechanisms that a ect charitable giving in traditional contexts are also likely to play a role in crowdfunding campaigns
(van Teunenbroek et al., 2022). Among these are altruism (Fehr & Fischbacher, 2003), the joy of giving (i.e., warm glow Andreoni, 1990) (i.e., warm glow) and solicitation (van Teunenbroek & Hasanefendic, 2023). van Teunenbroek and Dalla Chiesa (2022) also mention feeling part of a community as a motive. Project backers cannot only donate, but also comment on the project and share it on social media. Thus, by donating, donors can become part of a community (Josefy et al., 2017). The narrative of a project, communicated for example through the project description, can also motivate donors to contribute (van Teunenbroek & Dalla Chiesa, 2022). For example, Wang et al. (2022) found that a guilt evoking project description positively influenced the willingness to donate. Campaign pictures also influence potential donors. As mentioned by van Teunenbroek and Dalla Chiesa (2022), Rhue and Robert (2018) found that campaigns that depict people with happy facial expressions raised more money than campaigns depicting people with neutral facial expressions. van Teunenbroek and Dalla Chiesa (2022) note that the crowdfunding environment is characterized by high uncertainty because there exists information asymmetry between the donors and the project initiators. According to van Teunenbroek and Dalla Chiesa (2022), people use quality signals to guide their behavior in such situations (van Teunenbroek et al., 2020). Therefore, the perceived quality of a project is positively related to its funding success (Mollick, 2014). Similarly, the number of campaign updates is also positively related to campaign success (Mollick, 2014). As mentioned by van Teunenbroek and Dalla Chiesa (2022), the perceived credibility of the project initiator also

feeds into the perceived quality of a project. For example, Hörisch (2015) found that projects initiated by an o cially recognized non-profit organization tend to be more successful. With crowdfunding being an inherently online-based fundraising channel, social media plays an important role. Sharing a project on social media increases its visibility and therefore solicits potential donors (Bhati & McDonnell, 2020; Priante et al., 2022). Unsurprisingly, the number of social media shares is positively related to a project's success (Kubo et al., 2021).

van Teunenbroek et al. (2022) conducted a literature review about mechanisms that a ect giving via philanthropic crowdfunding. Based on the review, they developed a conceptual model that specifies how these mechanisms mediate the e ect of crowdfunding features (e.g., project description) on giving behavior. The crowdfunding features they study are the project creator, social information, project description and rewards. A few of the mediating mechanisms were already mentioned above, namely the perceived credibility of the project initiator, the perceived quality of the project, the emotional reaction elicited by the projects as well as the identification with a community. van Teunenbroek et al. (2022) also specify the strength of the tie a donor has with the project initiator as a mechanism. I do not summarize the reward-specific mechanisms since the crowdfunding platform studied in this study is not reward-based. Knowing how crowdfunding campaign factors a ect the success of a campaign, we can now go on to discuss the identification strategy that we use to identify the e ect of the perceived victim group size on the amount of funds raised by a project.

Methods Data and Identification Strategy

I use data downloaded from GoFundMe to test the compassion fade e ect in a real-life donation setting. Data from GoFundMe have been successfully used by researchers to study nonexperimenter solicited charitable contributions in a real-world setting (Sisco & Weber, 2019). In March 2022, I downloaded more than 60,000 fundraising projects from four countries (the US, the UK, Australia and Canada). When visiting GoFundMe.com, people see a grid of fundraising projects (see Figure A1 in the appendix). This grid displays the most important information for each project. One can see the location of the project, the project profile picture, the title of the project and the first few words of the description, the target amount to be raised, the funds already raised and when the last donation was made. Fundraising progress is visualized by a green progress bar. Although projects on GoFundMe have a target amount, GoFundMe follows a direct donation structure (van Teunenbroek et al., 2022) that allows initiators to keep the donated money, regardless of whether the target was reached or not. Raising more than the target amount is also possible. By clicking on a project from the project overview page, one is forwarded to the profile page of the project. On this page, project creators have the opportunity to display more photos and videos and to provide a detailed textual description of the project.

From the project overview page and the profile page, I download the data that I need to test hypothesis 1 (see Table 1). Each project belongs to a category (see Figure 1), and up to 1,000 projects per category can be downloaded.

My identification strategy relies on the backdoor criterion. A set of variables Z satisfies the backdoor criterion relative to an ordered pair of variables (x, y) if (1) Z blocks every path between x and y that contains an arrow into x and (2) no node in Z is a descendant of x. Given our treatment variable (number of persons on a project profile picture) and our dependent variable (amount of funds raised), I assume that the topic of a fundraising campaign, which I measure with the campaign description text and the campaign category, satisfies the backdoor criterion. As already mentioned, the topic of a campaign likely influences both how many people are shown on the project profile picture and how much people will donate to the project. The directed acyclic graph (DAG) (Rohrer, 2018) that visualizes this assumption is shown in Figure 2. We can directly condition on the category of a campaign since each campaign is assigned to a category. To control for the campaign description text, I use natural language processing (NLP) methods to convert the text data to a numerical representation. There is a relatively new but growing stream of literature on using textual data as controls in statistical models (Keith et al., 2020). As mentioned by (Keith et al., 2020), there exist multiple ways of measuring confounders from text, such as lexicons, supervised classifiers, topic models and embeddings. As I want to control for the overall topic of a project, I use the latter two methods. These methods inductively learn confounding factors to (ideally) account for all known and unknown aspects of the text (Keith et al., 2020). Topic models are generative probabilistic models that represent text as a mixture of latent topics (Roberts et al., 2014). Embeddings represent text as low-dimensional, dense vectors that encode the meaning of text. These numerical text representations are then used in place of the confounder (topic of the campaign) in a causal adjustment method (e.g., linear regression) (Keith et al., 2020). As our identification strategy crucially depends on our ability to measure the confounders from text, I use multiple text representation methods. Namely, I use topic models (Roberts et al., 2014) and two state-of-the-art document embedding techniques (Le & Mikolov, 2014; Reimers & Gurevych, 2019). This approach allows us to see how sensitive our estimates are to di erent text representations. I describe these methods in more detail below.

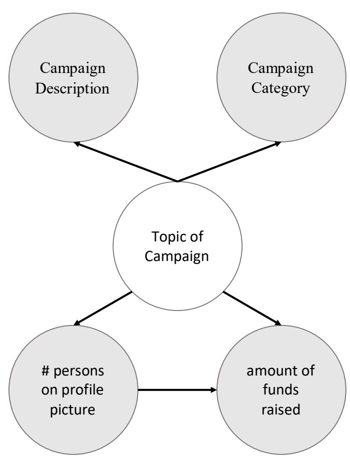


Figure 2

Directed acyclic graph (DAG) showing how I use the campaign category and campaign description to control for confounding between the number of persons on a project profile picture and the amount of funds raised by the project.

Control variables

As just mentioned, I use use the category of a fundraising project and the project description to identify the e ect of the perceived victim group size on the amount of funds raised (i.e., to control for confounding). However, there are other variables that, although they are not confounders, can increase the precision of the estimate of interest (Cinelli et al., 2022). The section on the "success factors of donation based crowdfunding campaigns" motivates and informs the selection of these variables. Although the e ects of these variables on campaign success were already discussed, I still briefly state the reason for inclusion when presenting the control variables. I include the total photos of a fundraising project, the number of updates posted and the length of the project description as control variables because studies have found that better documented projects raise more money (Wu et al., 2022). I also control for whether the fundraiser is organized by an organization or by people, whether it is a team fundraiser or not, and how many people are organizing the fundraiser. I include these controls because these variables likely a ect the sharing of the fundraiser, which increases the visibility of the fundraiser and thus likely also the amount of funds raised (Kubo et al., 2021). For the same reason I also created variables that control for whether the fundraiser was organized for anyone or not and if so whether it was organized for another person or to benefit an organization. I use the state-of-the-art named entity recognition model "ner-english" provided by the flair Python library to do this (Akbik et al., 2018).

To control for the popularity of the fundraiser, I include the number of times the fundraiser was shared on social media, the number of hearts (i.e., likes) the fundraiser collected and the number of comments that were made on the fundraiser project page as control variables. All of these variables likely a ect the visibility of the fundraiser, which in turn should a ect the amount of funds raised. For the same reason, I also control for the page position of the fundraiser in the category project overview page, as projects that appear on top of the page should receive more attention. Many of these control variables are also included to ensure that

I control for the number of people who directly visit a fundraising project without browsing other projects beforehand. This approach is crucial since my hypothesis rests on the assumption that potential donors are in a joint evaluation context when they decide on which project they should donate to. Controlling for variables that a ect the number of people who directly visit a fundraising project without browsing other projects before or after should ensure that donations that were made in a separate evaluation context do not a ect the results. The target amount is also included as a control as it has been shown to be associated with campaign success (Mollick, 2014). For reasons that are obvious I also control for the days that passed since the project launched.

Finally, I also control for the emotions displayed by the people who are depicted on the project profile photo. I do this because the facial expressions of victims have been shown to a ect giving behavior (Rhue & Robert, 2018; Small & Verrochi, 2009). Since this approach is only possible for projects that depict at least one face on the project profile page, I fit all models once without controlling for depicted facial emotions and once with controlling for depicted facial emotions.

When thinking about which controls to include, one must make sure that no "bad" controls are included (Cinelli et al., 2022). Colliders are an example of such bad controls. Conditioning on a collider, i.e., a common e ect of the exposure and outcome, leads to a noncausal association between the exposure and the outcome (Cinelli et al., 2022; Hünermund et al., 2021). In my case, what I call the "social" variables (number of social media shares, number of comments and number of campaign hearts) could potentially be such collider variables. People could share the fundraiser on social media, like or comment because they are depicted on the fundraiser’s project profile picture. Making a donation could also lead people to do these same things, which could make these variables serve as colliders between the number of people depicted on the project profile picture and the amount of funds raised. I therefore also fit the regression models without these three variables as controls (see appendix). However, I think that including these variables for the reasons stated above is more important, which is why these variables are included by default. However, the results for the models without these variables are very similar to those with these variables. In fact, the e ect of the number of persons on the project profile picture is even stronger when not controlling for these variables. Because the estimated e ects of these control variables are unlikely to have a causal interpretation, I follow previously made recommendations to not report them (Hünermund & Louw, 2020; Westreich & Greenland, 2013).

Natural Language Processing Methods

Embeddings embed text into a dense numerical space. There exist several methods to do this. These methods can roughly be divided into contextualized and noncontextualized methods. Word2vec (Mikolov et al., 2013) is a popular method used to obtain noncontextualized embeddings. Word2vec encodes words into numerical vectors by predicting a target word by its context words (or vice versa) with a shallow neural network. This prediction task is only a means to an end to obtain the weights of the neural network that are then used as the numerical vectors that represent a given target word. Using such a prediction task to obtain numerical representations has the advantage that words that are used in similar contexts end up having similar numerical representations (Camacho-Collados & Pilehvar, 2018). This process builds on the distributional hypothesis, which assumes that words that occur in similar contexts have similar meanings (Firth, 1957). After training word2vec on a (preferably large) corpus, one ends up with a vector of typically approximately 100-300 dimensions for each word of the corpus. To obtain a document vector, one can use the mean of all the word vectors that make up a document (Lau & Baldwin, 2016). Although word2vec uses the context of words to compute the word vectors, it does not assign di erent representations to the same word used in di erent contexts. For example, the word bank has the same numerical representation regardless of whether it is used in a financial context or not.

Contextualized models such as BERT (Devlin et al., 2018) alleviate this shortcoming by producing context-dependent embeddings for each word. These methods thus produce one word embedding for each unique word and each unique context the word appears in (Liu et al., 2020). Due to this approach, among other things, these models have achieved groundbreaking results in natural language understanding tasks (Rogers et al., 2020). Since BERT produces one embedding per word, we also need a way to aggregate those embeddings over the course of a document. I use Sentence BERT (SBERT) to do this (Reimers & Gurevych, 2019). Similar to averaging the word2vec embeddings per document, SBERT adds a pooling operation (mean) to the output of BERT to derive a fixed-sized document embedding. While contextualized models objectively perform better in most areas, they are not without limitations. First, because these models are very memory intensive, there is a limit to the length of text they can process. Second, since these methods are pretrained, they might fail to capture peculiarities of the text data at hand. I therefore use both of these methods to obtain embeddings of the project descriptions since word2vec can be trained on the data at hand and does not have a text length limit.

In addition to embeddings, I also use topic models to operationalize the topic of a campaign. Topic models are a popular method to detect latent topics in a collection of texts (Roberts et al., 2014). Topic models treat each document as a mixture of topics and each topic as a mixture of words. Latent Dirichlet allocation (LDA) is a popular method for fitting such topic models. I use the implementation provided by Roberts et al. (2014) to fit the topic model and use the method developed by Mimno and Lee (2014) to decide on the number of topics per topic model.

As evidenced by this short presentation of di erent methods for encoding the topic of a campaign from its campaign description text, these methods have di ering strengths and weaknesses. For example, while the strength of embeddings is that they promise to encode all aspects of a text (e.g., meaning, a ect, topic), their high level of dimensionality could complicate inference. The reverse is true for topic models; they only encode the topic(s) of the text but are often lower in dimension (i.e., number of topics) than embeddings. By using these di erent methods with di ering strengths and weaknesses, we can verify how sensitive our estimates are to the type of text encoding.

Inference Methods

To estimate the e ect of the number of persons on a campaign profile picture, I mainly rely on regression models and use double machine learning to assess the robustness of the results. I use double machine learning (Chernozhukov et al., 2017) as a robustness check because it allows us to control for confounders in a flexible (i.e., nonlinear) way.

Double machine learning can be illustrated with a partially linear model in the following form:

$$Y = D\Diamond_0 + g_0(X) + ', \quad \mathrm{E}(' \mid D,X) = 0, \tag{1}$$

$$D = m_0(X) + V, \quad \mathrm{E}(V \mid X) = 0, \tag{2}$$

where Y is the outcome variable (i.e., the amount of funds raised by a project) and D is the treatment variable (i.e., the number of persons on a projects profile picture). The (potentially) high-dimensional vector X contains the confounding and control variables, and $'$ and V are stochastic errors. Equation (1) is the equation of interest, and $\Diamond_0$ is the main regression coe cient that we would like to infer. Assuming that D is conditionally exogenous, $\Diamond_0$ has the interpretation of a structural or causal parameter. Equation (2) keeps track of confounding, i.e., the dependence of D on covariates. These covariates X a ect the treatment variable D via the function $m_0(X)$ and the outcome variable via the function $g_0(X)$.

Applying machine learning methods directly to Equations (1) and (2) may have a very high level of bias, which is caused by the regularization properties of machine learning algorithms (Bach et al., 2021; Chernozhukov et al., 2017). Double machine learning uses orthogonalization to overcome this regularization bias. To illustrate this, we rewrite the abovementioned PLR model in the following residualized form:

$$W = V \lozenge_0 + ', \qquad E(' \mid D, X) = 0, \tag{3}$$

$$W = (Y \neq l_0(X)), \qquad l_0(X) = E(Y \mid X) = 0, \tag{4}$$

$$V = (D \neq m_0(X)), \qquad m_0(X) = E(D \mid X) = 0, \tag{5}$$

Given identification, double machine learning then estimates l_0 and m_0 by $\hat{l}_0$ and $\hat{m}_0$ by solving the two problems of predicting Y and D using X. These prediction problems can be solved by using any o -the-shelf machine learning method. This gives us the following estimated residuals:

$$\hat{W} = Y \neq \hat{l}_0(X), \tag{6}$$

$$\hat{V} = D \neq \hat{m}_0(X). \tag{7}$$

To avoid overfitting, these residuals are of a cross-validated form. We can then finally estimate $\lozenge_0$ by regressing the residual $\hat{W}$ on $\hat{V}$. Conventional inference is used for this final regression estimator. Double machine learning uses a method-of-moments estimator for $\lozenge_0$ with a Neyman-orthogonal score function. This approach ensures that the moment condition used to identify and estimate $\lozenge_0$ is insensitive to small perturbations of the nuisance functions (i.e., $\hat{l}_0(X)$ and $\hat{m}_0(X)$) estimated by the machine learning models. Although this ensures some robustness, a good approximation of the nuisance functions is still crucial. I therefore use three di erent machine learning algorithms, namely, regression trees (Therneau et al., 2015), random forests (Breiman, 2001) and XGBoost (T. Chen & Guestrin, 2016). I refer interested readers to Chernozhukov et al. (2017) for a detailed treatment of double machine learning.

To ensure that the machine learning methods can well approximate the nuisance functions, I train the methods via random search (Bergstra & Bengio, 2012) for 20 iterations each. To set the tuning space of the hyperparameters, I rely on the current advice from the literature (Bischl et al., 2023).

Person and facial emotion detection

Given the large number of project profile pictures, it is not feasible to detect the number of persons in a project profile picture and the facial expressions shown by these people by hand. I therefore rely on machine learning algorithms for this task. The state-of-the-art algorithms for these tasks are as good or even better than human raters while being significantly faster. To detect the number of persons in a project profile picture, I use a Faster R-CNN model (Ren et al., 2015) that was pretrained to detect persons in the COCO dataset (Lin et al., 2014). I use the mmdetection Python library (K. Chen et al., 2019) to implement the model. More information regarding this model is provided in the appendix.

To detect the facial emotions expressed by the people on the project profile picture, I use a model that is considered state of the art in facial expression detection at the time of performing the analysis (Savchenko, 2021). I use the hsemotion Python library to detect expressed facial emotions with this model. For each detected face, this model returns the probability that this face shows emotion x for a total of seven emotions (angry, disgust, fear, happiness, sadness, surprise, and neutral). To obtain one value per project profile picture, I take the mean per emotion when multiple faces are detected.

Fitted models and preprocessing

To ensure that the results are not driven by outliers, I follow previous literature that used data from GoFundMe (Sisco & Weber, 2019) and removed projects that raised more than the mean plus 3 standard deviations per country. The analyses conducted on the full sample are reported in the appendix. The results for the models with outliers are similar to those without outliers but tend to have larger standard errors. To account for heteroscedasticity, heteroskedasticity-consistent standard errors (HC1) are used. To ensure comparability to Garinther et al. (2022), I run the analysis once for all projects and once only for projects that have twelve or fewer people on the project profile picture. This approach also increases the comparability with other studies; M.-R. Li and Yin (2022) found that in studies that showed participants pictures of beneficiaries, showing eight beneficiaries was the most frequent approach. Since studies conducted in the lab usually have at least one person in the solicitation picture, I also run the analysis for all projects and only for projects with at least one person in the project profile picture.

Results Descriptive Results

The data without outliers contain a total of 64,024 fundraising projects without any missing data in the variables listed in Table 1. This table also contains summary statistics for all of these variables. On average, these projects raised $4,390, with a standard deviation of $11,586. There are on average 2.6 persons depicted on the project profile picture, with a standard deviation of 5.5 persons. The majority of the projects (59.5%) have at least one person on the project profile picture. Only a few projects (4.8%) have more than twelve persons on the project profile picture. The majority of projects are thus comparable to the experimental literature with regard to the number of persons shown on the donation request (M.-R. Li & Yin, 2022).

The subset of projects with at least one detected face on the project profile picture contains 29,446 projects (Table 2). The mean facial emotions by project profile picture are mostly happy and neutral. The appendix shows the values given in Tables 1 and 2 for the full sample (i.e., with outliers). As I consider projects that raised more than the mean plus three standard deviations per country as outliers, 824 (1.27%) of the projects were excluded from the sample with all projects, and 464 (1.56%) projects were excluded from the sample where at least one face was detected on the project profile page.

E ect of perceived victim group size on funds raised

Figure 3 shows the e ect of the number of persons on a project's profile picture on the amount of funds raised by a project. For each country, text control method and subset of data, four models are fitted (three double machine learning models and one regression model). I first present the results that consider all projects (top row for each country panel). For projects in Australia, the number of people on the picture has no significant e ect on the amount of funds raised. For Canada, there is a small but significant e ect for most of the models. For the UK, most models indicate that there is no significant e ect. For US-based projects, most models indicate a positive e ect; however, compared to the **Table 1**

Variable	NotNA	Mean	Median	Sd	Min	Max
Number of persons on campaign picture	64024	2.596	1	5.46	0	84
At least one person on campaign picture	64024					
... No	25927	40.5%				
... Yes	38097	59.5%				
Max. 12 people on campaign picture	64024					
... No	3060	4.8%				
... Yes	60964	95.2%				
Amount raised	64024	4389.74	768.416	11585.774	0	137690
Target amount	64024	11344.629	3475.118	35071.549	0.695	1108890
Created x days ago	64024	63.625	45	78.796	1	3089
Length of description (words)	64024	205.693	152	208.625	0	8256
Total updates of campaign	64024	0.634	0	2.303	0	117
Total photos of campaign	64024	1.846	1	3.254	0	218
Number of social media shares	64024	95.81	11	305.226	0	18445
Number of campaign hearts	64024	46.253	17	164.025	0	21028
Number of comments	64024	1.901	0	9.252	0	1074
Organized by	64024					
... an organization	3590	5.6%				
... a person	60434	94.4%				
Number of people organizing	64024	1.117	1	1.464	0	138
Team fundraiser	64024					
... No	58913	92%				
... Yes	5111	8%				
Organized for	64024					
... not organized for anyone	50801	79.3%				
... an organization	4785	7.5%				
... a person	8438	13.2%				
Country	64024					

... Australia	16114	25.2%
... Canada	14574	22.8%
... UK	17004	26.6%
... USA	16332	25.5%

Summary Statistics of all projects without outliers

Table 2

Variable	NotNA	Mean	Median	Sd	Min	Max
Number of persons on campaign picture	29446	4.533	2	6.754	0	80
At least one person on campaign picture	29446					
... No	811	2.8%				
... Yes	28635	97.2%				
Max. 12 people on campaign picture	29446					
... No	2536	8.6%				
... Yes	26910	91.4%				
Mean facial emotion: angry	29446	0.062	0.019	0.113	0	0.99
Mean facial emotion: disgust	29446	0.048	0.015	0.096	0	0.981
Mean facial emotion: fear	29446	0.025	0.003	0.064	0	0.964
Mean facial emotion: happy	29446	0.559	0.608	0.371	0	1
Mean facial emotion: sad	29446	0.063	0.021	0.108	0	0.98
Mean facial emotion: surprise	29446	0.047	0.011	0.09	0	0.965
Mean facial emotion: neutral	29446	0.195	0.119	0.212	0	0.972
Amount raised	29446	3405.149	997.359	6919.25	0	72677
Target amount	29446	11562.584	3890.6	32217.524	0.695	800865
Created x days ago	29446	60.297	42	80.774	1	3089
Length of description (words)	29446	223.752	168	220.584	1	8256
Total updates of campaign	29446	0.69	0	2.413	0	117
Total photos of campaign	29446	1.878	1	3.424	0	218
Number of social media shares	29446	132.291	36	351.934	0	16599
Number of campaign hearts	29446	57.962	24	115.497	0	4972
Number of comments	29446	2.39	1	5.691	0	214
Organized by	29446					
... an organization	1229	4.2%				
... a person	28217	95.8%				
Number of people organizing	29446	1.132	1	1.273	0	105
Team fundraiser	29446					
... No	26871	91.3%				
... Yes	2575	8.7%				
Organized for	29446					
... not organized for anyone	22742	77.2%				
... an organization	1612	5.5%				
... a person	5092	17.3%				
Country	29446					

... Australia	6826	23.2%
... Canada	6397	21.7%
... UK	6989	23.7%
... USA	9234	31.4%

Summary Statistics of projects with at least one detected face without outliers

other countries, there is more variance in the estimates between the di erent models. In contrast to the double machine learning results, the regression results indicate no positive e ect in all but the topic model estimate.

Restricting the analysis to projects that show no more than twelve people on the project profile picture increases the e ect of the number of people on the picture on the amount of funds raised across all countries. For most countries and models, the estimates show that an additional person on the project profile picture leads to an increase in the amount of funds raised of approximately $25. The models presented in the last two rows for each country only consider projects with at least one person on the project profile picture, namely, once for all projects where this condition is met and once further restricted to those with a maximum of twelve people on the picture. The results are very similar to those models where all projects are considered. Overall, these results suggest that there is a small, but not consistent, e ect of the number of persons on the project profile picture on the amount of funds raised. The positive e ect is most consistent for the subset of projects with at most twelve people on the project profile picture. Most importantly, not one of the estimates of the 256 fitted models shown in Figure 3 indicates a significant negative e ect of the number of persons on the project profile picture on the amount of funds raised.

Figure 4 shows the same analysis as that shown in Figure 3, but it includes models that also account for the facial emotion expressed by the people on the project profile picture. The results are very similar to the results reported in Figure 3. The models that consider all projects within this subsample mostly indicate no significant e ect, except for projects posted in the US, where the majority of models indicate a positive e ect. As in Figure 3, the e ects become larger when further restricting the sample to projects with a maximum of twelve people on the project profile picture. Most models still indicate no e ect for projects posted in Australia and Canada. However, for the UK, there is now a consistent positive e ect and a mostly consistent positive e ect for the US.

The fact that models with projects that have at most twelve people on them show larger e ects could indicate a nonlinear (i.e., concave) e ect of the number of people on a project profile picture on the amount of funds raised. I therefore also fit the models shown in Figures 3 and 4 with a quadratic e ect of the number of persons on a project's profile picture. I only use ordinary least squares regression to fit these models because the results of the regression and double machine learning estimates are very similar.

Figure 5 shows the results of the same models as those fitted in Figure 3, but it includes an additional quadratic term for the number of persons on a project profile picture. In the majority of the models, the quadratic term has a small but significant negative e ect. Compared to Figure 3, the linear e ect of the number of persons on the project profile picture seems to be larger.

Looking at the models that also control for the emotions expressed by the people on the project profile picture (Figure 6), we see a similar result. In most of the models, the quadratic e ect is significantly negative. Except for projects from the UK, there is a significant positive linear e ect of the number of people on the project profile picture on the amount of funds raised. The e ect of an additional person on the project profile picture on the amount of funds raised is considerable (approximately \$100 for projects in Australia and Canada and around \$500 for projects in the US for models fitted on the subgroup of projects with a maximum of twelve people on the project profile picture).

Looking at the results with regards to hypothesis 1, I thus find that the postulated positive e ect of the perceived victim group size (i.e., number of persons on the project profile picture) on the amount of funds raised by the project is mixed. I found the most consistent positive e ect for models that were fitted on the subset of projects with a maximum of twelve people on the project profile picture. In the estimates obtained with the double machine learning estimators, projects in the UK and the US show the most consistent positive e ect. Interestingly, the e ect vanishes for projects from the UK in the models where a nonlinear e ect was included and the facial emotions were controlled for (Figure 6). However, all other models with this specification for the other three countries show a significant positive e ect of the number of people on the project profile picture on the amount of funds raised. I can thus confirm hypothesis 1 for all but projects from the UK.

Discussion

Crowdfunding has great promise both for individual and institutional fundraisers (Alexiou et al., 2020). However, most fundraising projects fail to reach their targets (Kenworthy & Igra, 2022). In this study I explored whether this lack of success could in part be explained by the di erent characteristics of crowdfunding and more traditional fundraising (e.g., mail solicitation). Among other things, crowdfunding platforms di er from more traditional means of fundraising in that potential donors are in a joint evaluation context, i.e., they can choose among a large number of projects to donate to. In contrast, traditional fundraising often occurs in a separate evaluation context, where potential donors often face only one donation request at a time. Importantly for fundraisers, laboratory studies have shown that the e ect of the victim group size on donations reverses when going from a separate to a joint evaluation context (Erlandsson, 2021; Garinther et al., 2022). In contrast to results obtained in a separate evaluation context, people in a joint evaluation context donate more to larger victim groups (Erlandsson, 2021; Garinther et al., 2022). Using data from over 60,000 GoFundMe crowdfunding projects from four countries, I tested whether this e ect generalizes to a real-world setting (i.e., crowdfunding). I did this by testing the e ect of the number of people on a project's profile picture (i.e., perceived victim group size) on the amount of funds raised by a project. Consistent with recent evidence from the lab that tested this e ect in a joint evaluation condition (Garinther et al., 2022), I found a mostly significant positive e ect of the number of people depicted on the project profile picture on the amount of funds raised by the project for the subgroup of projects that is most similar to the stimuli used in the laboratory (i.e., maximum twelve people on the picture). The results of the models fitted on the full sample were more mixed, and only a few of them indicated a significant positive e ect. This discrepancy calls for more experimental studies that vary the range of victims beyond what has mostly been done in the literature (e.g., a maximum of twelve victims).

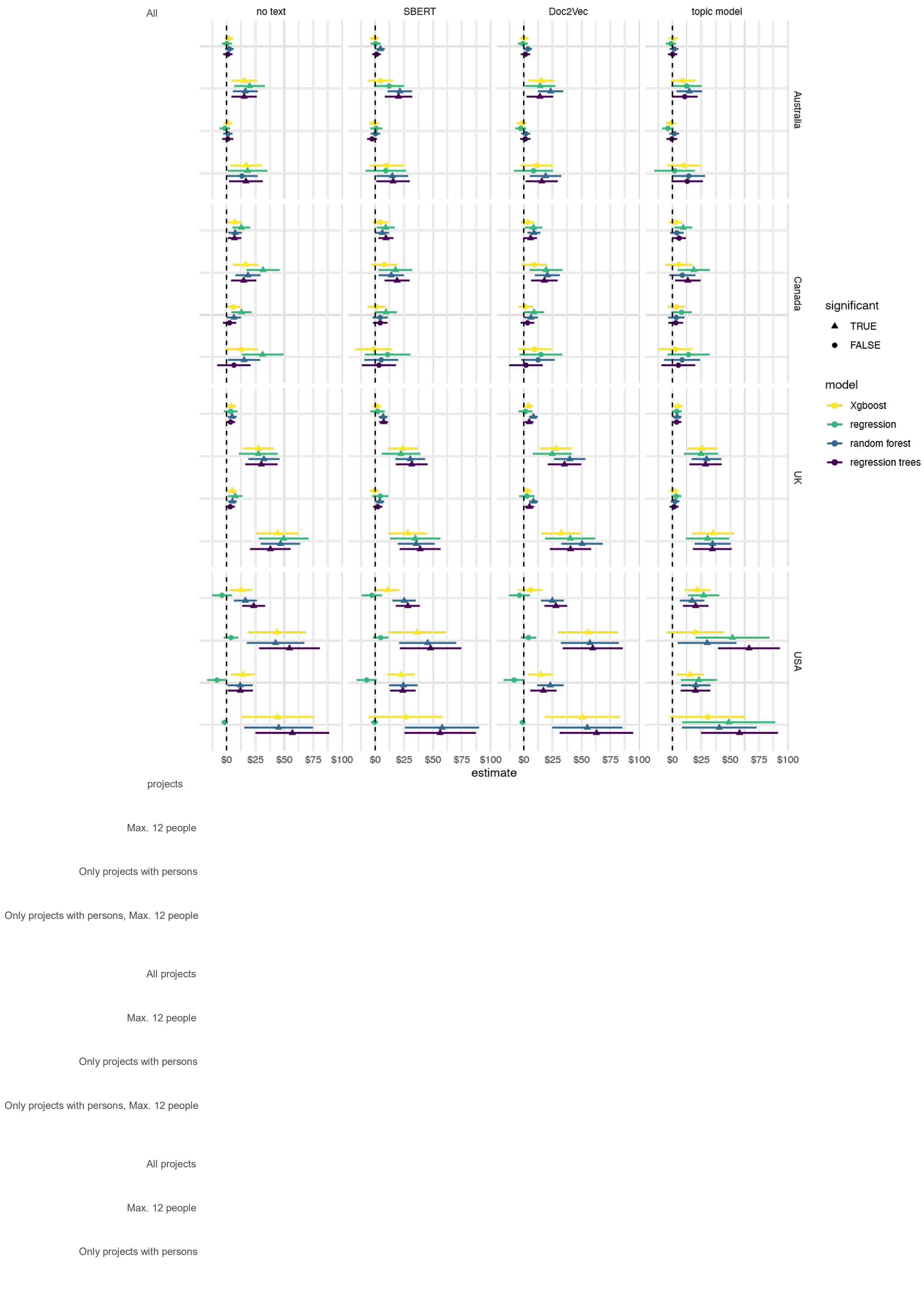
Most of the models that include the quadratic e ect indicate a concave e ect of the perceived victim group size on the amount of funds raised. This has not yet been found in experimental studies. However, the nonlinear e ect is rather small; therefore, laboratory studies might lack the power to detect this e ect. This nonlinear e ect might indicate that even in a joint evaluation setting, people might be prone to a ective biases. The a ective bias perspective denotes that in separate evaluations, people's numeracy limitations and biases in a ective processing (Hamilton & Sherman, 1996; Slovic, 2007) might be responsible for the compassion fade e ect (Butts et al., 2019). My results show that these biases could still a ect decision-making in a joint evaluation setting but are trumped by attributes that have a high level of justifiability (i.e., victim group size) (Erlandsson, 2021). Thus, to attract more donations from people who are browsing projects on GoFundMe, it is still beneficial to include more rather than fewer people on the project profile picture. This advice is relative to the other projects that share the same category (e.g., medical, sports). As shown in Figure 1, the variance of the number of people depicted on the profile picture is rather small within projects that share the same category. My results hold for the typical range of depicted persons per category and do not necessarily extrapolate beyond that range. I thus advise people who want to raise money in a joint evaluation context to increase the victim group size by a sensible amount. For example, when raising money for a sick child, show the whole family (but not, e.g., the whole child's school class) instead of only the child. Assuming that people use the (perceived) victim group size (high justifiability) (Erlandsson, 2021) to choose from among similar fundraising projects but are also a ected by a ective biases to some extent (Butts et al., 2019), having a marginally higher victim size than that of other similar projects should be most e ective. The assumptions used to derive this advice are consistent with my results and with the evidence quoted by Erlandsson (2021) that shows that emotional reactions are more predictive of attitudes toward policies in separate evaluations (Ritov & Baron, 2011), while e ciency-related attributes are more predictive in joint evaluations (Bazerman et al., 2011; Caviola et al., 2014). Future research could test whether there are other fundraising related attributes that have di ering e ects depending on whether people evaluate them in a separate or joint evaluation context. For example, the overhead ratio is an e ciency-related attribute that is di cult to evaluate in separate evaluation and should thus receive more weight in a joint evaluation context.

My results are not without limitations. First, as with any observational study aiming to draw causal conclusions, my results crucially depend on the identifying assumptions. I use the project category and the project description to control for confounders between the number of people shown on the project profile picture and the amount of funds raised. I also use a number of other control variables that should help with identifying the variance needed to draw causal conclusions. However, it is possible that I have omitted a confounder or included a bad control. While this possibility cannot be ruled out, the fact that the results are consistent with results from the laboratory and that the results also hold for the models without the (potentially) bad controls indicates that the results are hopefully not a ected by such a possibility. Second, I rely heavily on machine learning algorithms to conduct this study. While these algorithms are already very good, the field is developing rapidly. Methods that use text data to adjust for confounding are constantly evolving (Feder et al., 2022). While I use three di erent methods to control for the topic of a fundraiser (i.e., campaign description), these methods will soon be surpassed by newer and better methods. The third limitation concerns the data. While it covers countries from three di erent continents, my analysis is still restricted to a Western sample. Due to cultural di erences in how crowdfunding projects are set up (Cho & Kim, 2017), my results do not necessarily generalize to, e.g., more collectivist cultures (Nie et al., 2022). Future research could replicate my results in such cultures.

While experimental studies are immune to some of these limitations, there is only so much we can learn about real-world behavior from laboratory studies (Levitt & List, 2007). Experiments can provide evidence for a causal relationship between X and Y within the constraints of the experimental design, but they are often unable to shed light on the strength of an e ect in everyday natural settings (Diener et al., 2022). As argued by Diener et al. (2022), a research program is most successful when experiments are integrated with other methods rather than being considered the sole source of valid information. I follow this perspective and use evidence and theories derived from laboratory studies as the basis for this study to test whether those findings are generalizable to the field. Even though the results from these laboratory studies have direct and easy-to-implement implications for fundraisers, field studies that test whether these e ects are actually generalizable are still missing from the literature. This is in contrast to the broader fundraising literature, where field studies are not uncommon (e.g., Alston et al., 2021; Woods et al., 2023). The lack of evidence from field data applies to psychological research more generally (Diener et al., 2022; Grosz et al., 2020). I hope that this study helps to alleviate this shortcoming by showing the usefulness of observational field data to complement and extend findings from the laboratory.

Notes

1. Child sponsorships, such as those used by, e.g., World Vision International, are a notable exemption.


All
no text
SBERT
Doc2Vec
topic model
Australia
Canada
UK
USA
significant
TRUE
FALSE
model
Xgboost
regression
random forest
regression trees
$0 $25 $50 $75 $100
estimate
projects
Max. 12 people
Only projects with persons
Only projects with persons, Max. 12 people
All projects
Max. 12 people
Only projects with persons
Only projects with persons, Max. 12 people
All projects
Max. 12 people
Only projects with persons
Only projects with persons, Max. 12 people

All projects

Max. 12 people

Only projects with persons

Only projects with persons, Max. 12 people

Figure 3

E ect of the number of persons on a project’s profile picture on the amount of funds raised. Error bars denote 95% confidence intervals.

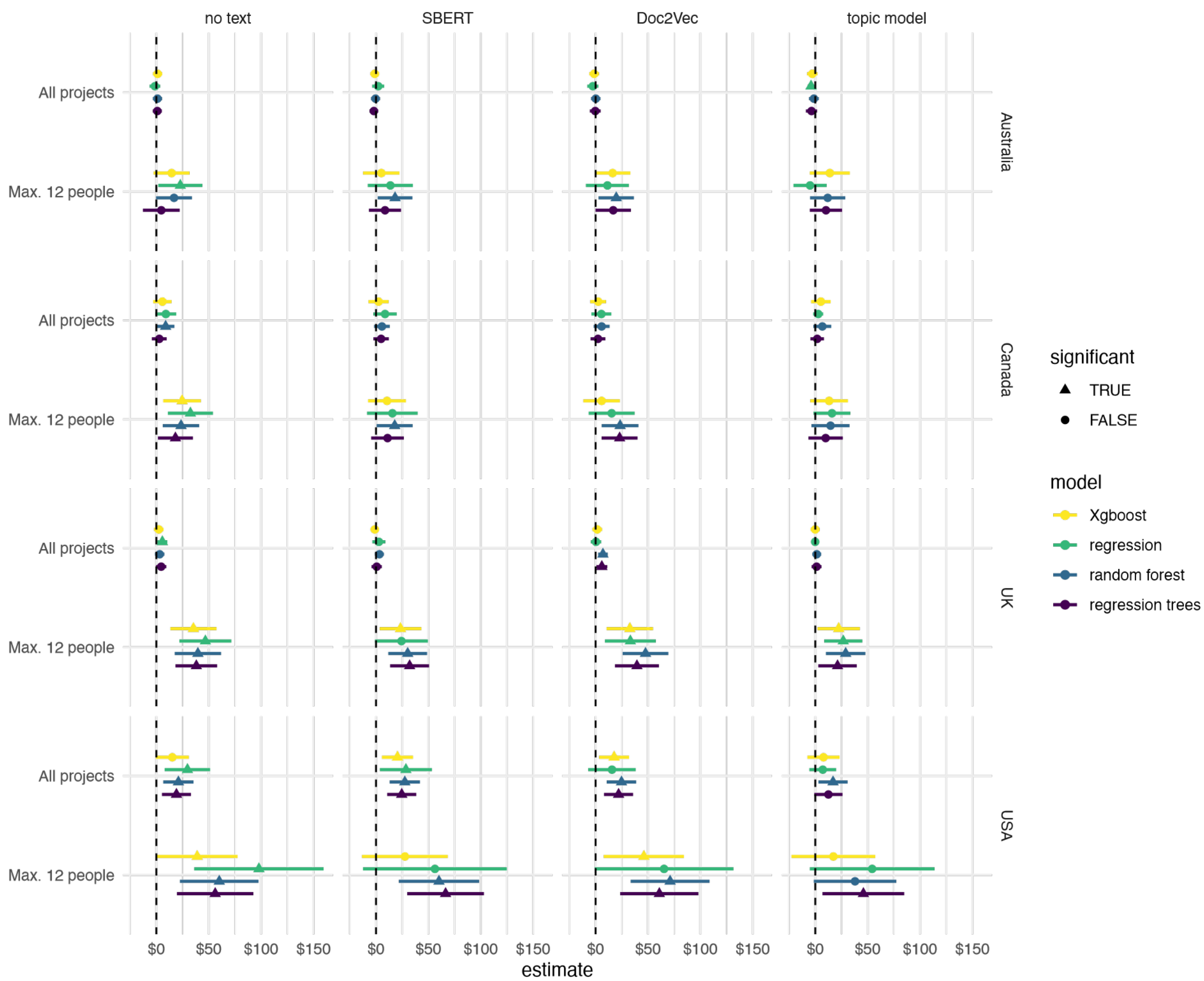


Figure 4

E ect of the number of persons on a project’s profile picture on the amount of funds raised for models that control for the facial emotions expressed by the persons. Error bars denote 95% confidence intervals.

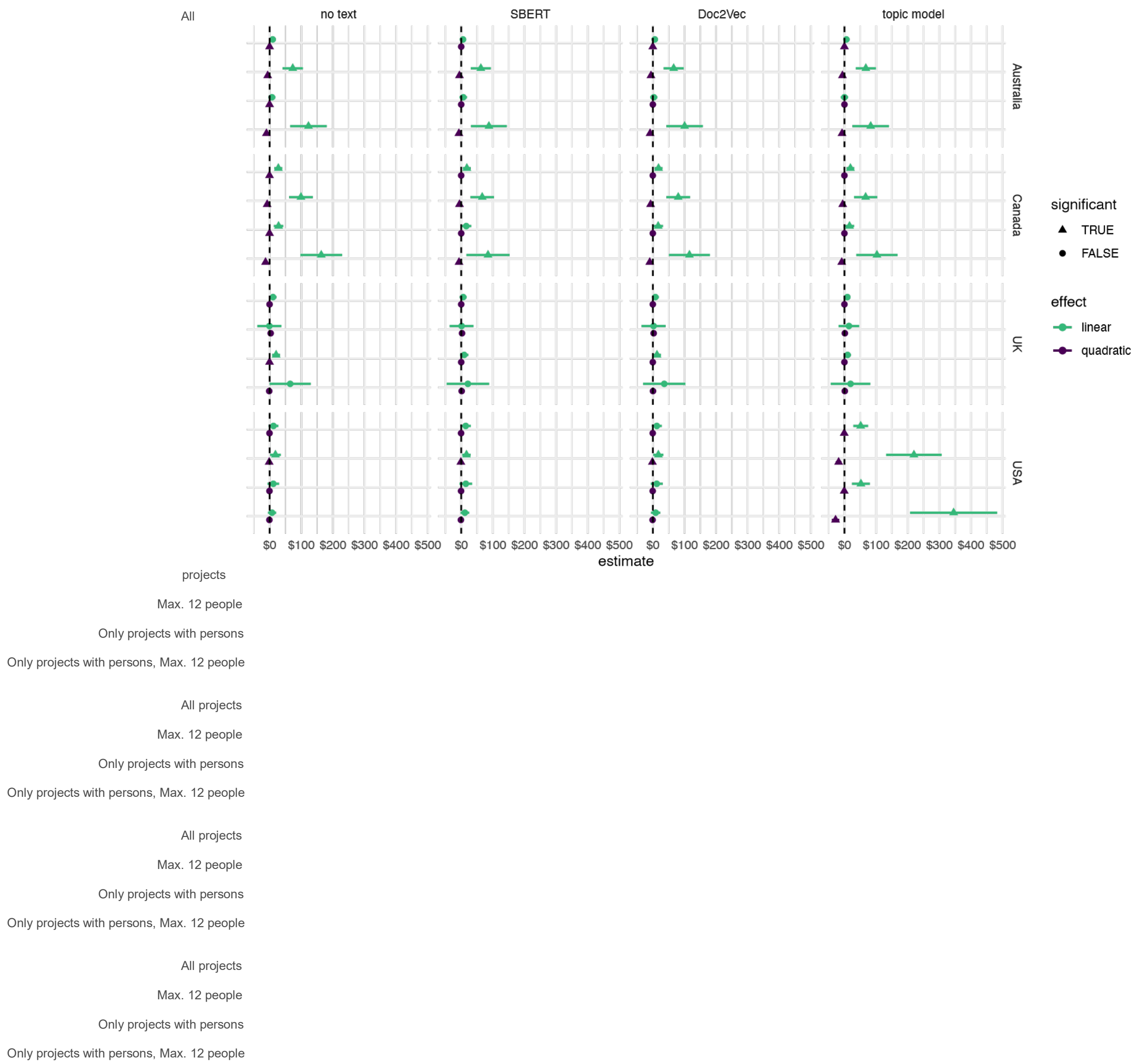


Figure 5

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised. Error bars denote 95% confidence intervals.

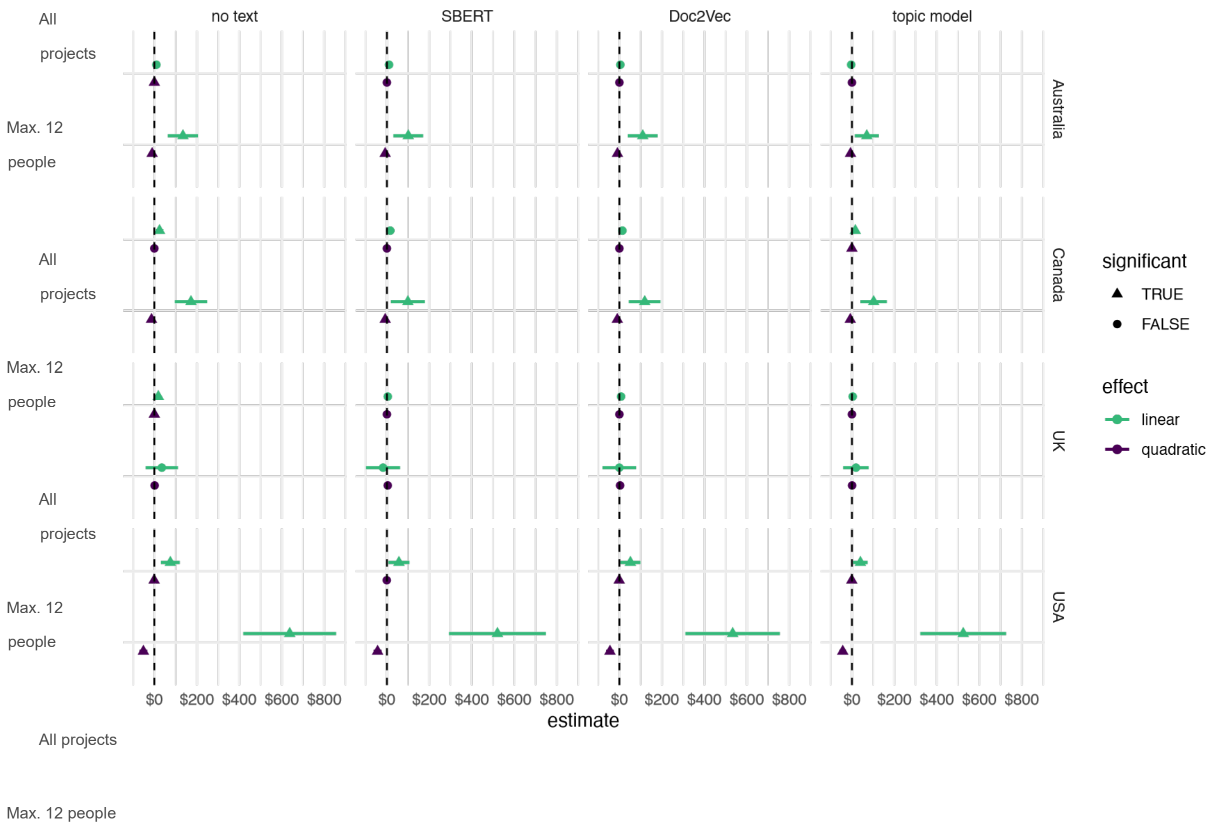


Figure 6

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised for models that do control for the facial emotions expressed by the persons. Error bars denote 95% confidence intervals.

References

Akbik, A., Blythe, D., & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. *COLING 2018, 27th International Conference on Computational Linguistics*, 1638–1649.

Alexiou, K., Wiggins, J., & Preece, S. B. (2020). Crowdfunding acts as a funding substitute and a legitimating signal for nonprofit performing arts organizations. *Nonprofit and Voluntary Sector Quarterly*, *49*(4), 827–848. https://doi.org/10.1177/0899764020908338

Alston, M., Eckel, C., Meer, J., & Zhan, W. (2021). High-capacity donors' preferences for charitable giving. *Nonprofit and Voluntary Sector Quarterly*, *50*(6), 1262–1273.

https://doi.org/10.1177/08997640211003251

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, *100*(401), 464–477. https://doi.org/https://doi.org/10.2307/2234133

Bach, P., Chernozhukov, V., Kurz, M. S., & Spindler, M. (2021). DoubleML–An Object-Oriented Implementation of Double Machine Learning in R. *arXiv preprint arXiv:2103.09603*. https://doi.org/https://doi.org/10.48550/arXiv.2103.09603

Bazerman, M. H., Gino, F., Shu, L. L., & Tsay, C.-J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emotion review*, *3*(3), 290–292. https://doi.org/10.1177/1754073911402370

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bhati, A., & McDonnell, D. (2020). Success in an online giving day: The role of social media in fundraising. *Nonprofit and Voluntary Sector Quarterly*, *49*(1), 74–92. https://doi.org/10.1177/0899764019868849

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, e1484. https://doi.org/https://doi.org/10.1002/widm.1484

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/https://doi.org/10.1145/2133806.2133826

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? a theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16–33.

https://doi.org/https://doi.org/10.1016/j.obhdp.2018.12.006

Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, *63*, 743–788. https://doi.org/https://doi.org/10.1613/jair.1.11259

Cameron, C. D. (2017). Compassion collapse: Why we are numb to numbers? *The oxford handbook of compassion science*, *261*, 271.

Cameron, C. D., & Payne, B. K. (2011). Escaping a ect: How motivated emotion regulation creates insensitivity to mass su ering. *Journal of personality and social psychology*, *100*(1), 1. https://doi.org/https://doi.org/10.1037/a0021643

Caviola, L., Faulmüller, N., Everett, J. A., Savulescu, J., & Kahane, G. (2014). The evaluability bias in charitable giving: Saving administration costs or saving lives? *Judgment and decision making*, *9*(4), 303–315. https://doi.org/https://doi.org/10.1017/S1930297500006185

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al. (2019). Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*. https://doi.org/https://doi.org/10.48550/arXiv.1906.07155

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794. https://doi.org/https://doi.org/10.1145/2939672.2939785

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment e ects. *American Economic Review*, *107*(5), 261–65. https://doi.org/10.1257/aer.p20171038

Cho, M., & Kim, G. (2017). A cross-cultural comparative analysis of crowdfunding projects in the united states and south korea. *Computers in Human Behavior*, *72*, 312–320. https://doi.org/https://doi.org/10.1016/j.chb.2017.03.013

Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls.

Sociological Methods & Research, 00491241221099552. https://doi.org/10.1177/00491241221099552

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/https://doi.org/10.48550/arXiv.1810.04805

Dickert, S., & Slovic, P. (2009). Attentional mechanisms in the generation of sympathy. *Judgment and Decision making*, *4*(4), 297–306. https://doi.org/https://doi.org/10.1017/S1930297500003879

Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, *17*(4), 1101–1119. https://doi.org/https://doi.org/10.1177/17456916211037670

Dovidio, J. F., Piliavin, J. A., Gaertner, S. L., Schroeder, D. A., & Clark III, R. D. (1991). The arousal: Cost-reward model and the process of intervention: A review of the evidence. Sage Publications, Inc.

Erlandsson, A. (2021). Seven (weak and strong) helping e ects systematically tested in separate evaluation, joint evaluation and forced choice. *Judgment and decision making*, *16*(5), 1113–1154. https://doi.org/https://doi.org/10.1017/S1930297500008378

Feder, A., Keith, K. A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., Eisenstein, J., Grimmer, J., Reichart, R., Roberts, M. E., et al. (2022). Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics*, *10*, 1138–1158. https://doi.org/https://doi.org/10.1162/tacl_a_00511

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791. https://doi.org/https://doi.org/10.1038/nature02043

Figley, C. R. (1995). Compassion fatigue: Toward a new understanding of the costs of caring. The Sidran Press.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Garinther, A., Arrow, H., & Razavi, P. (2022). Victim number e ects in charitable giving: Joint evaluations promote egalitarian decisions. *Personality and Social Psychology Bulletin*, *48*(1), 34–48. https://doi.org/https://doi.org/10.1177/0146167220982734

GoFundMe. (2022). *Gofundme's 2021 giving report*. Retrieved December 20, 2022, from https://www.gofundme.com/c/gofundme-giving-report-2021

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255. https://doi.org/https://doi.org/10.1177/1745691620921521

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological review*, *103*(2), 336.

Hart, T. R. (2002). Ephilanthropy: Using the internet to build support. *International Journal of Nonprofit and Voluntary Sector Marketing*, *7*(4), 353–360. https://doi.org/https://doi.org/10.1002/nvsm.192

Hörisch, J. (2015). Crowdfunding for environmental ventures: An empirical analysis of the influence of environmental orientation on the success of crowdfunding initiatives. *Journal of cleaner production*, *107*, 636–645. https://doi.org/https://doi.org/10.1016/j.jclepro.2015.05.046

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational behavior and human decision processes*, *67*(3), 247–257. https://doi.org/https://doi.org/10.1006/obhd.1996.0077

Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological science*, *24*(9), 1801–1808. https://doi.org/10.1177/0956797613482947

Hsee, C. K., Zhang, J., Wang, L., & Zhang, S. (2013). Magnitude, time, and risk di er similarly between joint and single evaluations. *Journal of Consumer Research*, *40*(1), 172–184. https://doi.org/https://doi.org/10.1086/669484

Hünermund, P., & Louw, B. (2020). On the nuisance of control variables in regression analysis. *arXiv preprint arXiv:2005.10314*. https://doi.org/https://doi.org/10.48550/arXiv.2005.10314

Hünermund, P., Louw, B., & Caspi, I. (2021). Double machine learning and automated confounder selection – a cautionary tale. *arXiv preprint arXiv:2108.11294*. https://doi.org/10.48550/ARXIV.2108.11294

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim e ect. *Journal of Risk and uncertainty*, *14*(3), 235–257. https://doi.org/https://doi.org/10.1023/A:1007740225484

Josefy, M., Dean, T. J., Albert, L. S., & Fitza, M. A. (2017). The role of community in crowdfunding success: Evidence on cultural attributes in funding campaigns to "save the local theater". *Entrepreneurship Theory and Practice*, *41*(2), 161–182.

Keith, K. A., Jensen, D., & O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. *arXiv preprint arXiv:2005.00649*. https://doi.org/https://doi.org/10.48550/arXiv.2005.00649 Kenworthy, N., & Igra, M. (2022). Medical crowdfunding and disparities in health care access in the united states, 2016–2020. *American Journal of Public Health*, *112*(3), 491–498. https://doi.org/https://doi.org/10.2105/AJPH.2021.306617

Kogut, T., & Ritov, I. (2005a). The "identified victim" e ect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/https://doi.org/10.1002/bdm.492

Kogut, T., & Ritov, I. (2005b). The singularity e ect of identified victims in separate and joint evaluations. *Organizational behavior and human decision processes*, *97*(2), 106–116. https://doi.org/https://doi.org/10.1016/j.obhdp.2005.02.003

Kubo, T., Verıssimo, D., Uryu, S., Mieno, T., & MacMillan, D. (2021). What determines the success and failure of environmental crowdfunding? *Ambio*, *50*, 1659–1669.

https://doi.org/https://doi.org/10.1007/s13280-021-01522-0

Kuo, Y.-F., Lin, C. S., & Liu, L.-T. (2022). The e ects of framing messages and cause-related marketing on backing intentions in reward-based crowdfunding. *Journal of Retailing and Consumer Services*, *64*, 102799. https://doi.org/https://doi.org/10.1016/j.jretconser.2021.102799

Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368*. https://doi.org/https://doi.org/10.48550/arXiv.1607.05368

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *International conference on machine learning*, 1188–1196.

Lee, S., & Feeley, T. H. (2016). The identifiable victim e ect: A meta-analytic review. *Social Influence*, *11*(3), 199–215. https://doi.org/https://doi.org/10.1080/15534510.2016.1216891

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, *21*(2), 153–174. https://doi.org/10.1257/jep.21.2.153

Li, M.-R., & Yin, C.-Y. (2022). Facial expressions of beneficiaries and donation intentions of potential donors: E ects of the number of beneficiaries in charity advertising. *Journal of Retailing and Consumer Services*, *66*, 102915. https://doi.org/https://doi.org/10.1016/j.jretconser.2022.102915

Li, X., & Hsee, C. K. (2019). Beyond preference reversal: Distinguishing justifiability from evaluability in joint versus single evaluations. *Organizational Behavior and Human Decision Processes*, *153*, 63–74. https://doi.org/https://doi.org/10.1016/j.obhdp.2019.04.007

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European conference on computer vision*, 740–755.

https://doi.org/https://doi.org/10.1007/978-3-319-10602-1_48

Liu, Q., Kusner, M. J., & Blunsom, P. (2020). A survey on contextual embeddings. *arXiv preprint arXiv:2003.07278*. https://doi.org/https://doi.org/10.48550/arXiv.2003.07278

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). E cient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. https://doi.org/https://doi.org/10.48550/arXiv.1301.3781

Mimno, D., & Lee, M. (2014). Low-dimensional embeddings for interpretable anchor-based topic inference. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1319–1328.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, *29*(1), 1–16. https://doi.org/https://doi.org/10.1016/j.jbusvent.2013.06.005

Nie, L., Chan, K. N., & Lam, W.-F. (2022). Social expectations for charitable giving in china. *Nonprofit and Voluntary Sector Quarterly*, *0*(0), 08997640221100721. https://doi.org/10.1177/08997640221100721

Priante, A., Ehrenhard, M. L., van den Broek, T., Need, A., & Hiemstra, D. (2022). "mo" together or alone? investigating the role of fundraisers' networks in online peer-to-peer fundraising. *Nonprofit and Voluntary Sector Quarterly*, *51*(5), 986–1009. https://doi.org/10.1177/08997640211057456

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*. https://doi.org/https://doi.org/10.48550/arXiv.1908.10084

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Rhue, L., & Robert, L. P. (2018). Emotional delivery in pro-social crowdfunding success. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing*

Systems, 1–6. https://doi.org/https://doi.org/10.1145/3170427.3188534

Ritov, I., & Baron, J. (2011). Joint presentation reduces the e ect of emotion on evaluation of public actions. *Cognition and Emotion*, *25*(4), 657–675. https://doi.org/https://doi.org/10.1080/02699931.2010.512512

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, *58*(4), 1064–1082. https://doi.org/https://doi.org/10.1111/ajps.12103

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866. https://doi.org/https://doi.org/10.1162/tacl_a_00349

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in methods and practices in psychological science*, *1*(1), 27–42. https://doi.org/https://doi.org/10.1177/2515245917745629

Ruehle, R. C., Engelen, B., & Archer, A. (2021). Nudging charitable giving: What (if anything) is wrong with it? *Nonprofit and Voluntary Sector Quarterly*, *50*(2), 353–371. https://doi.org/10.1177/0899764020954266

Savchenko, A. V. (2021). Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)*, 119–124. https://doi.org/https://doi.org/10.1109/SISY52375.2021.9582508

Shaw, L. L., Batson, C. D., & Todd, R. M. (1994). Empathy avoidance: Forestalling feeling for another in order to escape the motivational consequences. *Journal of Personality and Social Psychology*, *67*(5), 879. https://doi.org/https://psycnet.apa.org/doi/10.1037/0022-3514.67.5.879

Sisco, M. R., & Weber, E. U. (2019). Examining charitable giving in real-world online donations. *Nature communications*, *10*(1), 1–8.

https://doi.org/https://doi.org/10.1038/s41467-019-11852-z

Slovic, P. (2007). If i look at the mass i will never act: Psychic numbing and genocide. *Judgment and Decision Making*, *2*(2), 79–95. https://doi.org/https://doi.org/10.1017/S1930297500000061

Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of marketing research*, *46*(6), 777–787. https://doi.org/https://psycnet.apa.org/doi/10.1509/jmkr.46.6.777

Therneau, T., Atkinson, B., & Ripley, B. (2015). Rpart: Recursive partitioning and regression trees. *R package version*, *4*, 1–9.

van Teunenbroek, C., & Dalla Chiesa, C. (2022). Motives for giving via crowdfunding in eight categories: A review and theoretical framework of donation- and reward-based crowdfunding motives. *International Crowdfunding Research Conference*, 1. van Teunenbroek, C., Bekkers, R., & Beersma, B. (2020). Look to others before you leap: A systematic literature review of social information e ects on donation amounts.

Nonprofit and Voluntary Sector Quarterly, *49*(1), 53–73. https://doi.org/https://doi.org/10.1177/0899764019869537

van Teunenbroek, C., Dalla Chiesa, C., & Hesse, L. (2022). The contribution of crowdfunding for philanthropy: A review and theoretical framework of reward- and donation-based crowdfunding.

van Teunenbroek, C., & Hasanefendic, S. (2023). Researching the crowd: Implications on philanthropic crowdfunding and donor characteristics during a pandemic. *Journal of Philanthropy and Marketing*, *28*(1), e1773. https://doi.org/https://doi.org/10.1002/nvsm.1773

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: A ect and charity are greatest for a single child in need. *PloS one*, *9*(6), e100115. https://doi.org/https://doi.org/10.1371/journal.pone.0100115

Wang, X., Zhang, X., & Tong, A. (2022). The impact of linguistic style of medical crowdfunding philanthropic appeals on individual giving. *Procedia Computer Science*, *199*, 293–300. https://doi.org/https://doi.org/10.1016/j.procs.2022.01.036

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coe cients. *American journal of epidemiology*, *177*(4), 292–298. https://doi.org/https://doi.org/10.1093/aje/kws412

Woods, A. L., Wu, F. Y., & Hebl, M. R. (2023). Giving to matthew, emily, jose, or maria: A field study examining the impact of race and gender on donation requests. *Nonprofit and Voluntary Sector Quarterly*, *0*(0), 08997640221140314. https://doi.org/10.1177/08997640221140314

Wu, Y., Zhang, X., & Xiao, Q. (2022). Appeal to the head and heart: The persuasive e ects of medical crowdfunding charitable appeals on willingness to donate. *Information Processing & Management*, *59*(1), 102792. https://doi.org/https://doi.org/10.1016/j.ipm.2021.102792

Appendix

Detailed information regarding machine learning algorithms

To detect the number of persons on a project's profile picture, I used the "faster_rcnn_r50_ca e_fpn_mstrain_3x_coco" model provided by the mmdetection Python library. This is a Faster R-CNN model (Ren et al., 2015) that was pretrained to detect persons in the COCO dataset (Lin et al., 2014) The model can be downloaded with this link. To detect the facial emotions expressed by the people on the project profile picture, I used the hsemotion Python library (Savchenko, 2021). I used the "enet_b2_7" model provided by this library.

Main results with outliers

Table A1

Summary Statistics of all projects

Variable	NotNA	Mean	Median	Sd	Min	Max
Number of persons on campaign picture	64848	2.597	1	5.445	0	84
At least one person on campaign picture	64848					
... No	26152	40.3%				
... Yes	38696	59.7%				
Max. 12 people on campaign picture	64848					
... No	3088	4.8%				
... Yes	61760	95.2%				
Amount raised	64848	6168.056	800	25684.513	0	1157925.116
Target amount	64848	13094.568	3475.118	41662.061	0.695	1108890
Created x days ago	64848	63.823	45	79.571	1	3089
Length of description (words)	64848	207.501	153	210.819	0	8256
Total updates of campaign	64848	0.657	0	2.382	0	117
Total photos of campaign	64848	1.877	1	3.394	0	218
Number of social media shares	64848	112.604	12	881.456	0	182965
Number of campaign hearts	64848	55.907	17	268.53	0	23577
Number of comments	64848	2.273	0	12.422	0	1074
Organized by	64848					
... organized by an organization	3660	5.6%				
... organized by a person	61188	94.4%				
Number of people organizing	64848	1.122	1	1.475	0	138
Team fundraiser	64848					
... No	59571	91.9%				
... Yes	5277	8.1%				
Organized for	64848					
... not organized for anyone	51221	79%				
... organized for an organization	4864	7.5%				
... organized for a person	8763	13.5%				
Country	64848					
... Australia	16325	25.2%				
... Canada	14757	22.8%				
... UK	17098	26.4%				
... USA	16668	25.7%				

Table A2

Variable	NotNA	Mean	Median	Sd	Min	Max
Number of persons on campaign picture	29910	4.516	2	6.716	0	80
At least one person on campaign picture	29910					
... No	819	2.7%				
... Yes	29091	97.3%				
Max. 12 people on campaign picture	29910					

... No	2550	8.5%				
... Yes	27360	91.5%				
Mean facial emotion: angry	29910	0.062	0.019	0.113	0	0.99
Mean facial emotion: disgust	29910	0.048	0.015	0.096	0	0.981
Mean facial emotion: fear	29910	0.025	0.003	0.064	0	0.964
Mean facial emotion: happy	29910	0.56	0.611	0.371	0	1
Mean facial emotion: sad	29910	0.063	0.021	0.108	0	0.98
Mean facial emotion: surprise	29910	0.047	0.011	0.09	0	0.965
Mean facial emotion: neutral	29910	0.194	0.118	0.211	0	0.972
Amount raised	29910	4714.578	1034.9	15801.836	0	787831
Target amount	29910	12871.291	3890.6	36190.548	0.695	800865
Created x days ago	29910	60.59	42	80.805	1	3089
Length of description (words)	29910	225.428	169	221.303	1	8256
Total updates of campaign	29910	0.723	0	2.495	0	117
Total photos of campaign	29910	1.916	1	3.55	0	218
Number of social media shares	29910	165.059	38	1238.569	0	182965
Number of campaign hearts	29910	76.528	24	294.934	0	23577
Number of comments	29910	3.117	1	12.7	0	720
Organized by	29910					
... organized by an organization	1240	4.1%				
... organized by a person	28670	95.9%				
Number of people organizing	29910	1.14	1	1.302	0	105
Team fundraiser	29910					
... No	27230	91%				
... Yes	2680	9%				
Organized for	29910					
... not organized for anyone	22901	76.6%				
... organized for an organization	1642	5.5%				
... organized for a person	5367	17.9%				
Country	29910					
... Australia	6965	23.3%				
... Canada	6486	21.7%				
... UK	7065	23.6%				
... USA	9394	31.4%				

Summary Statistics of projects with at least one detected face

Results of models that do not control for the number of social media shares, number of comments and number of campaign hearts

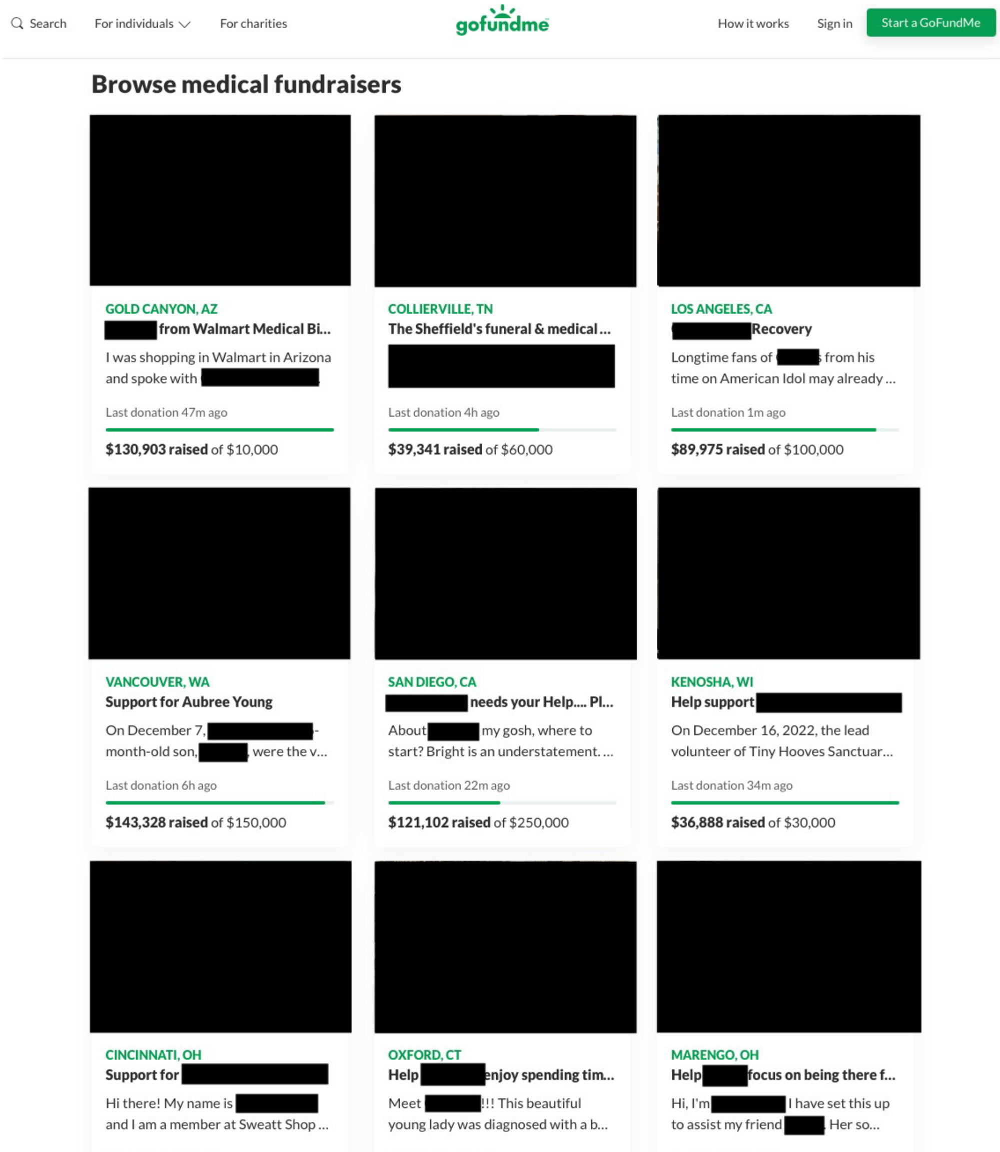
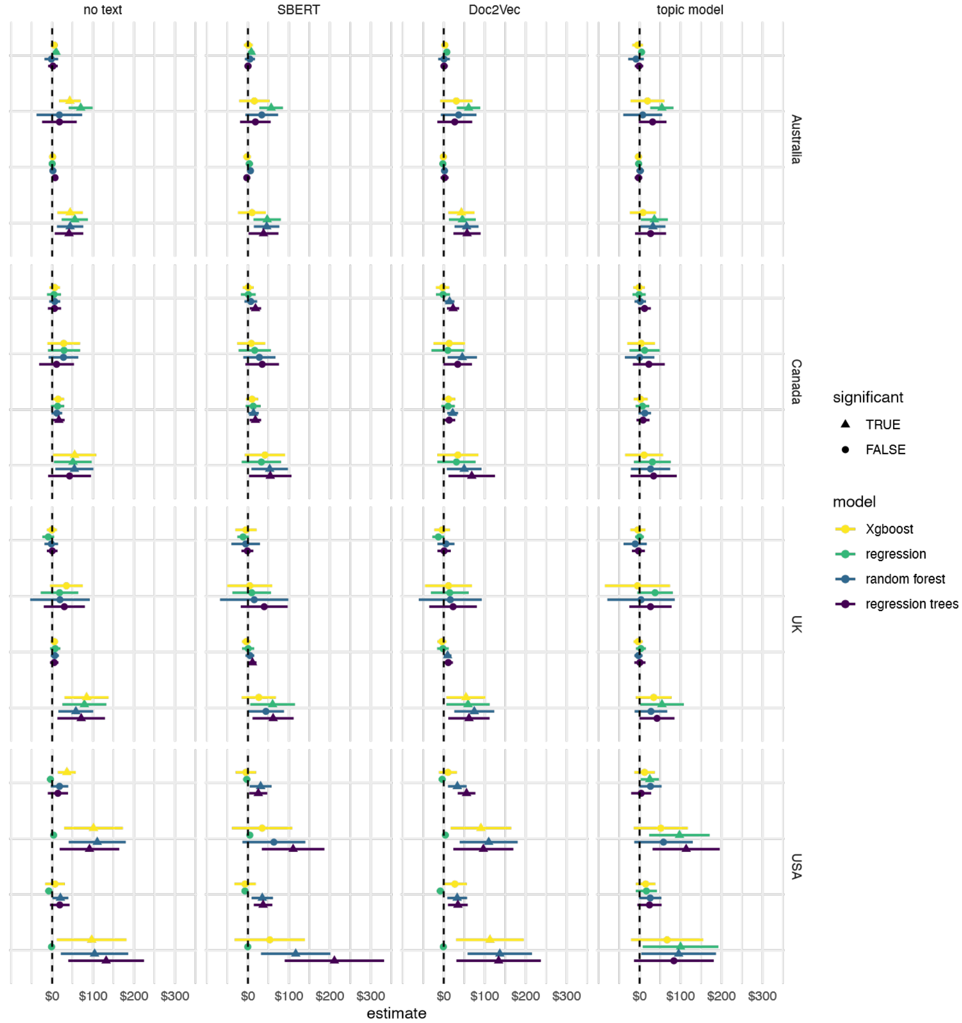


Figure A1

Project overview page for medical projects on GoFundMe.com. Identifying information blacked out.

All



projects

Max. 12 people

Only projects with persons

Only projects with persons, Max. 12 people

All projects

Max. 12 people

Only projects with persons

Only projects with persons, Max. 12 people

All projects

Max. 12 people

Only projects with persons

Only projects with persons, Max. 12 people

All projects

Max. 12 people

Only projects with persons

Only projects with persons, Max. 12 people

Figure A2

E ect of the number of persons on a project’s profile picture on the amount of funds raised in the full sample (i.e., with outliers). Error bars denote 95% confidence intervals.

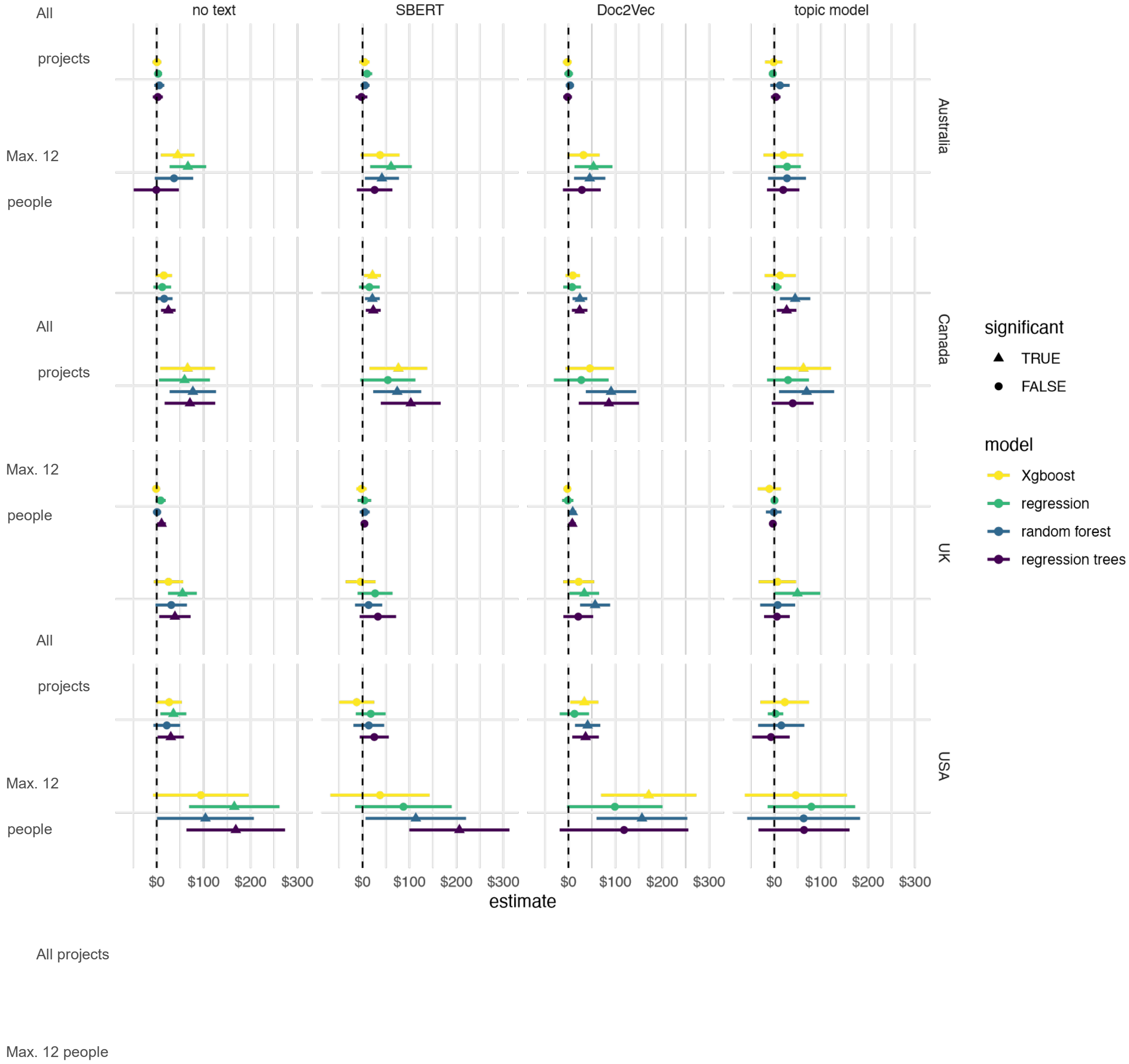


Figure A3

E ect of the number of persons on a project’s profile picture on the amount of funds raised for models that do control for the facial emotions expressed by the persons in the full sample (i.e., with outliers). Error bars denote 95% confidence intervals

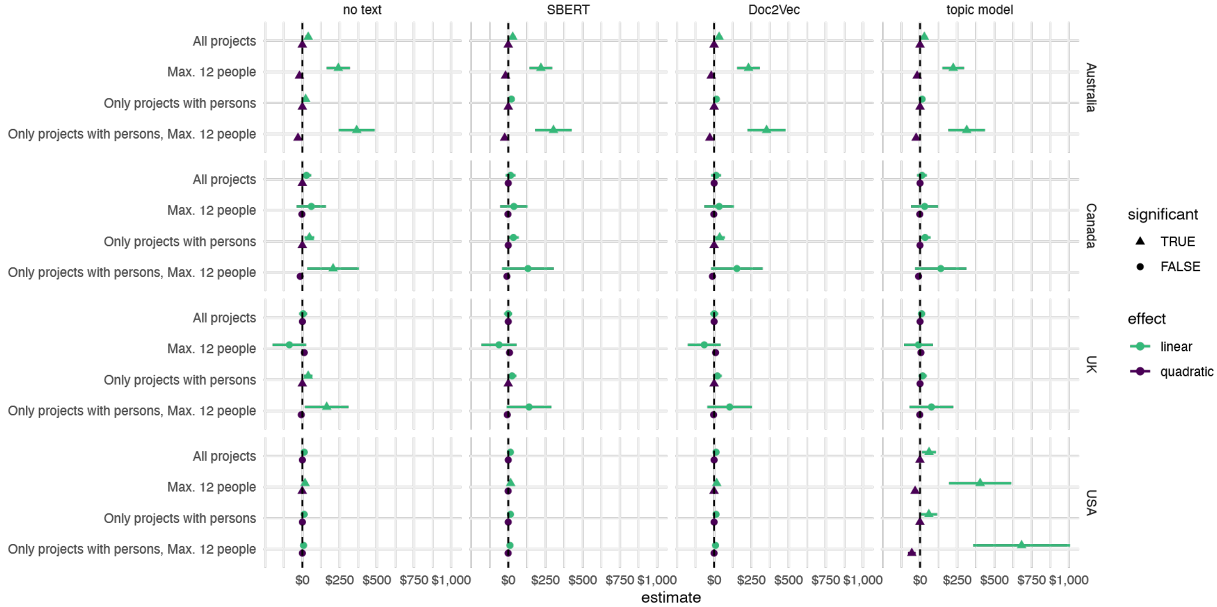


Figure A4

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised in the full sample (i.e., with outliers). Error bars denote 95% confidence intervals.

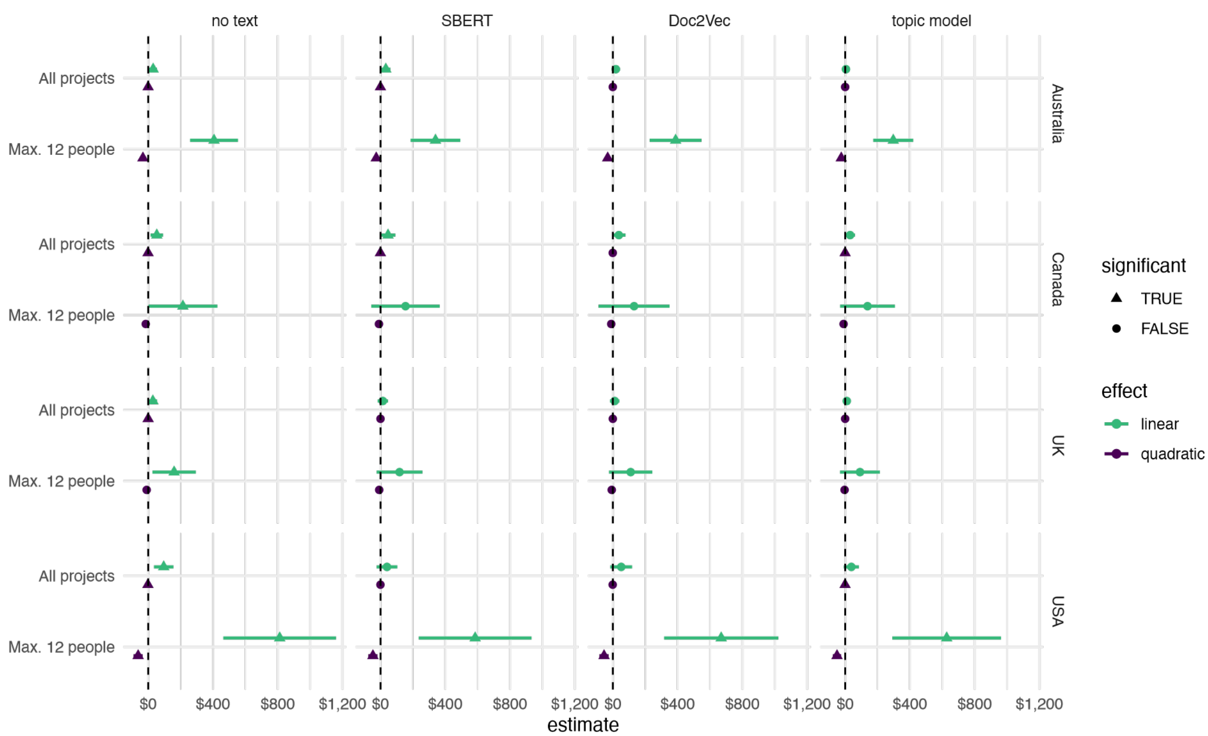


Figure A5

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised for models that do control for the facial emotions expressed by the persons in the full sample (i.e., with outliers). Error bars denote 95% confidence intervals.

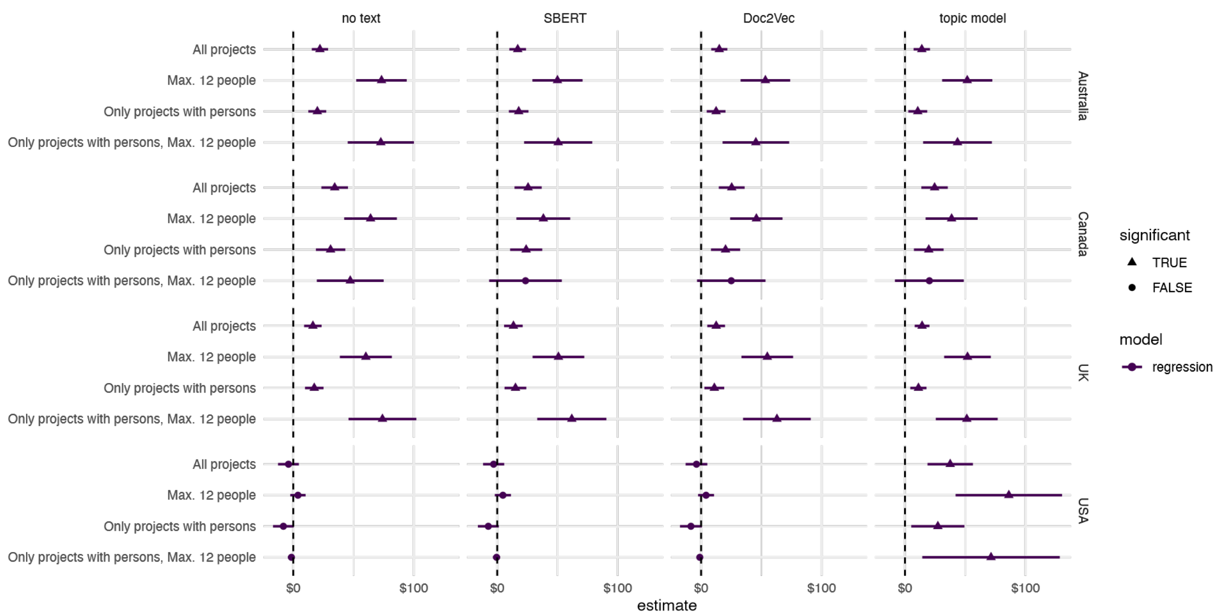


Figure A6

E ect of the number of persons on a project's profile picture on the amount of funds raised for models that do not control for social media shares, number of comments and number of campaign hearts. Error bars denote 95% confidence intervals.

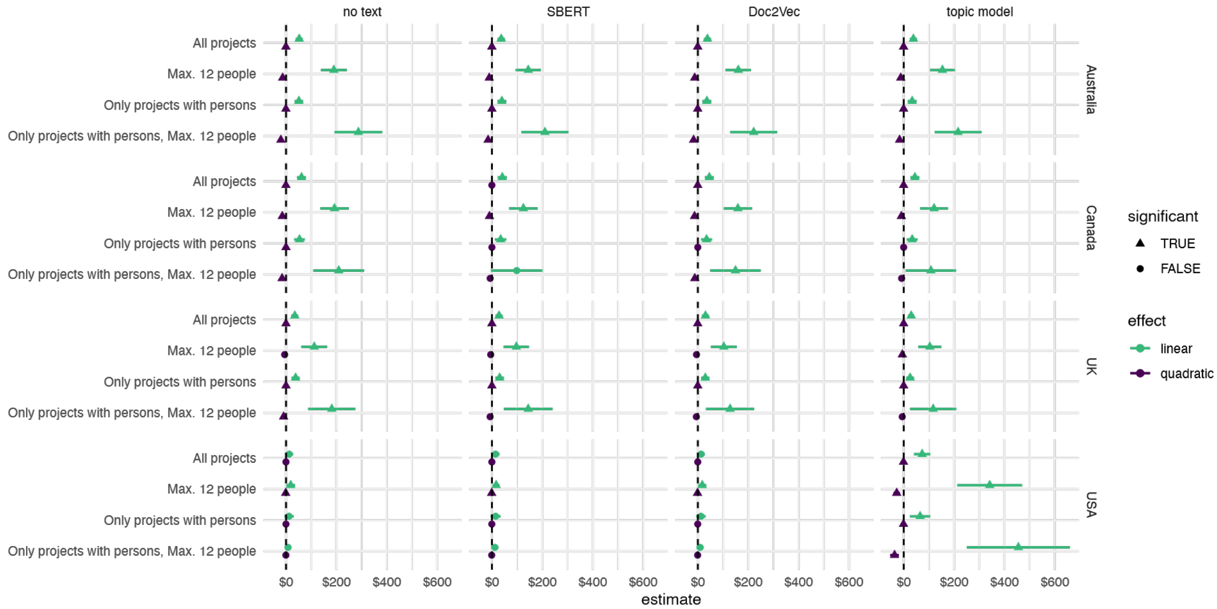


Figure A7

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised for models that do not control for social media shares, number of comments and number of campaign hearts. Error bars denote 95% confidence intervals.

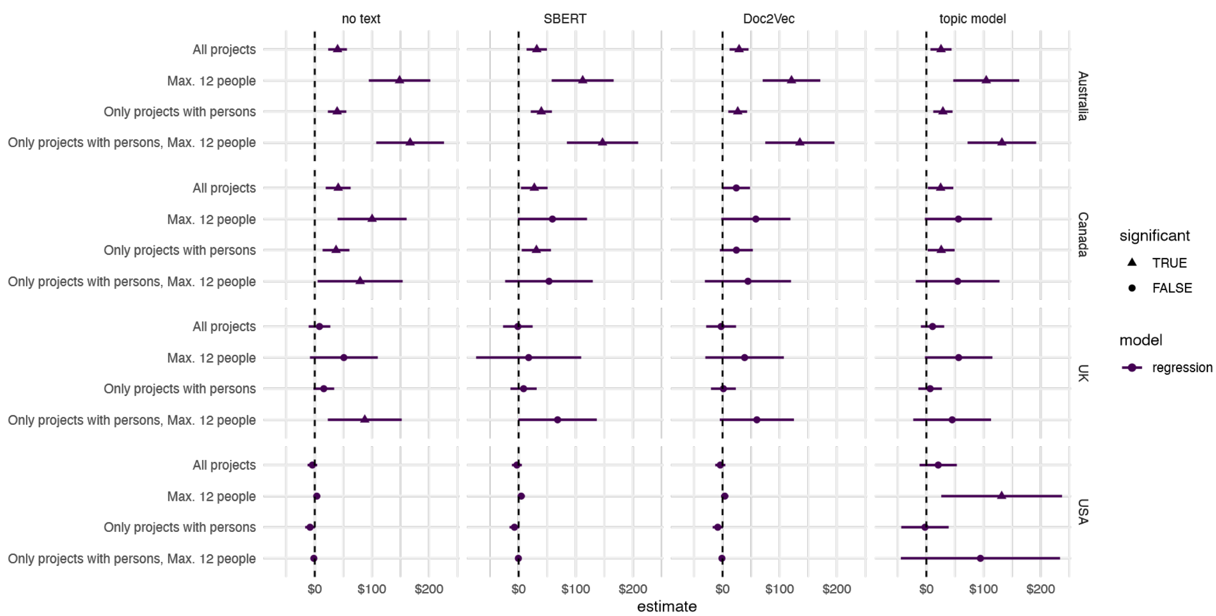


Figure A8

E ect of the number of persons on a project's profile picture on the amount of funds raised in the full sample (i.e., with outliers) for models that do not control for social media shares, number of comments and number of campaign hearts. Error bars denote 95% confidence intervals.

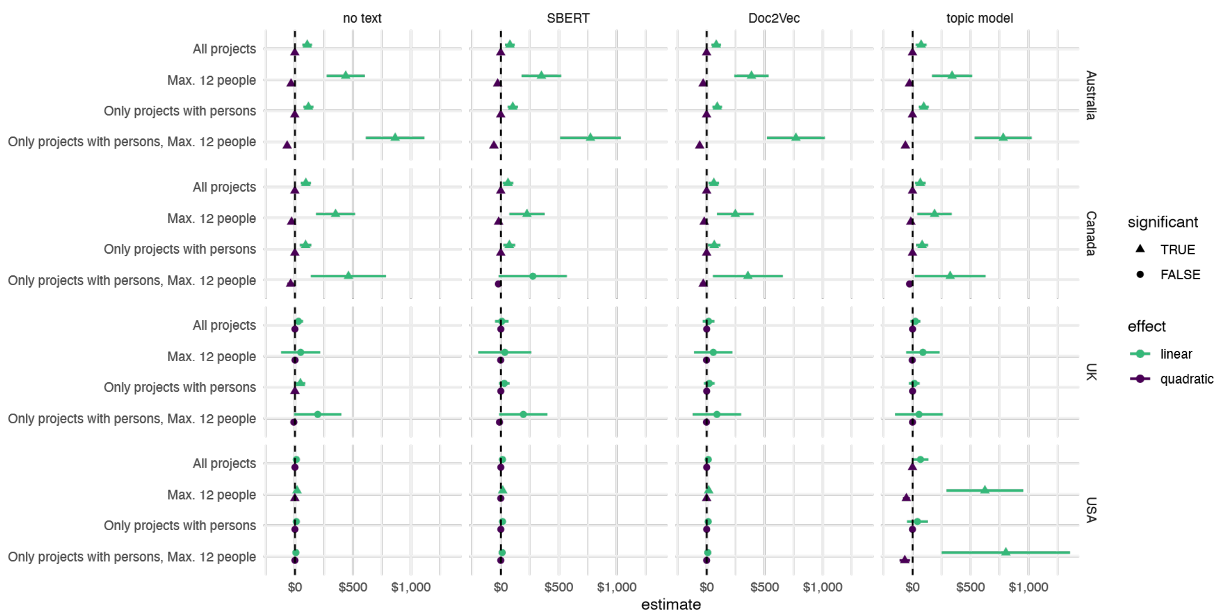


Figure A9

Linear and nonlinear e ect of the number of persons on a project's profile picture on the amount of funds raised in the full sample (i.e., with outliers) for models that do not control for social media shares, number of comments and number of campaign hearts. Error bars denote 95% confidence intervals.

Risking Your Health to Help Others: Volunteering During the COVID-19 Pandemic

Dominik S. Meier[1] [ORCID: 0000-0002-3999-1388], Amadeus Petrig[2], Georg von Schnurbein[1] [ORCID: 0000-0002-6660-1875]

[1]University of Basel, Switzerland

[2] Genossenschaft Migros Aare, Switzerland

Corresponding Author: Dominik S. Meier

Affiliations

Dominik S. Meier, Center for Philanthropy Studies and Center for Economic Psychology, University of Basel, Steinengraben 22, 4051 Basel, Switzerland, Email: dominik.meier@unibas.ch

Amadeus Petrig, Genossenschaft Migros Aare, Schanzenstrasse 5, 3000 Bern, Switzerland, Email: amadeus.petrig@migrosaare.ch

Georg von Schnurbein, Center for Philanthropy Studies, University of Basel, Steinengraben 22, 4051 Basel, Switzerland, Email: georg.vonschnurbein@unibas.ch

Author Biographies

Dominik S. Meier is a PhD student at the Center for Philanthropy Studies and the Center for Economic Psychology at the University of Basel (Switzerland). His research is focused on prosocial behavior and philanthropy.

Amadeus Petrig earned a Master of Science in Work and Organizational Psychology from the University of Fribourg (Switzerland). He works in the Venture Builder department at Migros, where he spearheads growth and innovation projects. During the COVID-19 crisis, he initiated and led the largest neighborhood project in Switzerland, AMIGOS.

Georg von Schnurbein is full professor for Foundation Management and director of the Center for Philanthropy Studies at the University of Basel (Switzerland). His research interests are nonprofit governance, nonprofit finance, and philanthropy.

Data availability statement

The data are not publicly available due to ethical, legal, or other concerns. Data can be made available upon request.

Acknowledgments

We would like to thank Peter Frumkin, Matthew Bennett, and all the participants of the 2022 Penn Summer Social Impact Doctoral Fellows Program for their valuable feedback on an earlier version of this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract

The COVID-19 pandemic affected the provision of voluntary work across the globe. We study informal volunteers who buy and deliver groceries for people in a high-risk group or in quarantine. Using data from a volunteering grocery delivering app in Switzerland that coordinated these volunteers, we are able to track volunteering during the pandemic. Combined with public health data on cases and deaths, we test how the severity of the pandemic affects the provision of voluntary work in the form of neighborhood grocery deliveries. We find a positive effect of the number of deaths on voluntary deliveries. However, in contrast to the literature studying the effect of the severity of the pandemic on giving, this effect is concave. We suggest that this concave effect is due to the signal of risk of infection implied by rising death rates, which is at odds with the signal of need to help others.

Keywords

informal volunteering, crisis, altruism, social solidarity, risk perception, volunteer coordination

Introduction

Catastrophes like earthquakes or pandemics can affect thousands of people, which can lead to a drastic increase in the demand for help. As catastrophes often occur unexpectedly, professional services might not be able to meet this demand for help (NHS, 2020), which often leads to volunteers joining the helping efforts. For example, survivors of an earthquake might join rescue efforts, and in the Covid-19 pandemic volunteers delivered groceries to people who had to self-isolate, as professional delivery services were not able to meet the demand (Meyersohn, 2020). While there are accounts showing that catastrophes lead to a wave of solidarity (Carlsen et al., 2020; Zaki, 2020), the jury is still out for how COVID-19 affected volunteering around the globe. While some spoke of a wave of solidarity that led to a surge in volunteering (UNRIC Brussels), other studies found that the pandemic led to a decrease in formal volunteering (Dederichs, 2022). While the effect of catastrophes like COVID-19 on helping behavior is still disputed, we know even less about how the severity of a catastrophe affects helping behavior. There is some evidence from cross-sectional studies showing that earthquake severity correlates with volunteer turnout (Iizuka & Aldrich, 2022). However, to the best of our knowledge, there is no evidence on how the severity of a long-lasting catastrophe like the Covid-19 pandemic affects individual helping behavior in the form of informal volunteering. From a theory standpoint, it is not clear whether the evidence from these cross-sectional studies translates to a longitudinal setting. From a standpoint of practical relevance, lack of such evidence makes disaster management suboptimal, since officials cannot gauge how professional helping services are complemented by informal help as the catastrophe intensifies or diminishes.

This work aims to provide such evidence by answering how the severity of the Covid-19 pandemic (as measured by the case- and death numbers) affects informal volunteering[6] (in the form of grocery food deliveries). To answer this research question, we use data on volunteer grocery food deliveries from a mobile application (app) that was designed to match people who need groceries delivered (i.e., because they had to self-isolate) with people who were willing to do so on a voluntary basis. The app launched in March 2020 as the first wave hit Switzerland and terminated its service in May 2021 as the need for this kind of help eventually subsided. Over

[6] Based on Cnaan et al. (1996), we broadly define volunteering as an activity that is performed to benefit others, is done out of free choice and is renumerated below the value of work provided. In contrast to formal volunteering, informal volunteering takes place outside of the organizational context (Brudney et al., 2019)

this time period, almost 27,000 people registered to do deliveries. The platform registered a total of 78,961 orders and 72,379 successful deliveries. We use fixed effects regression models to test how the weekly case- and death numbers relate to the number of weekly deliveries made by individuals. To identify the effect of the case- and death numbers on deliveries, we control for the number of orders placed and the number of other volunteers that are registered in a volunteer's delivery area.

We find that the death numbers have a concave (i.e., significant positive linear and negative quadratic) effect on the number of deliveries made. The case numbers show no consistent effect across the models. Thus, consistent with previous studies on how the severity of a catastrophe affects volunteering, the severity of the Covid-19 pandemic (i.e., number of deaths) seems to have had a positive effect on informal volunteering. However, previous studies reported a linear and not a concave effect (Izuka & Aldrich, 2022; Rotolo et al., 2015). We argue that the concavity of the effect is caused by the risk of getting infected when volunteers go out to deliver

groceries. Our results are relevant both from a theoretical and from a practical point of view. First, we show that the evidence from cross-sectional studies translates only partially to a longitudinal setting. Second, we show that the risk that volunteering poses to one's health seems to have a negative effect on informal volunteering. Third, our results are valuable for practitioners, as it allows for better planning on how volunteer helping services will complement professional helping efforts. As most orders were also delivered, our results also show that platform mediated matching of volunteers with volunteering opportunities is an effective and efficient way of volunteer management and might be a viable solution to deal with the problem of an oversupply of volunteers (Simsa et al., 2019). The following section gives a short overview of volunteering during the COVID-19 pandemic and then introduces the limited literature that examines how the severity of a disaster affects prosocial behavior (e.g., volunteering and donating). Section 2 introduces the methods and data used, section 3 presents the results, and section 4 discusses how the study contributes to research on informal crisis volunteering.

Volunteering during the COVID-19 pandemic

Only few studies provide insights regarding who engaged in (in)formal volunteering during the pandemic. Mak and Fancourt (2021) found that older people were more likely to participate in neighborhood volunteering than younger people. They also found that people who lived in urban areas were less likely to engage in neighborhood volunteering. Regarding psychosocial factors, they found that people with a larger social network and with higher levels of social support were more likely to participate in all types of voluntary work. People with a diagnosed disability or illness had lower odds of volunteering in neighborhood support (Mak & Fancourt, 2021). According to a review by Mao et al. (2021) local knowledge, social trust and social networks were key dimensions associated with community organizing and volunteering.

Volunteers were mostly women, middle-class, highly educated, and working-aged people. Similarly, Dederichs (2022) found that formal volunteering was more common among women, university graduates, elderly individuals, and those with high levels of self-rated health. These are characteristics that were also found to be positively related to volunteering before the pandemic (Wilson, 2012). Dederichs (2022) also investigated which socio-demographic characteristics are linked to formal volunteering in response to COVID-19. Being healthier, holding a university degree and being a woman was significantly positively associated with volunteering in response to Covid-19 (Dederichs, 2022). Already volunteering before the pandemic was the strongest predictor of volunteering in response to Covid-19 while age and whether children were present in the household had no significant effect.

Regarding the activities the volunteers engaged in, a review by Mao et al. (2021) suggests that especially during the early days of the pandemic, food shopping and providing emotional support were the most common activities. In the later stages of the pandemic, there was a shift towards activities that address the wider impact of the pandemic on areas such as homelessness, mental health, and employment (Mao et al., 2021). Jones et al. (2020) report that neighborhood support, from giving food and medical prescription assistance to providing health information and raising morale through humour was common in the UK. Mao et al. (2021) report a shift from offline to online volunteering, as the circumstances imposed by Covid-19 made offline volunteering challenging.

Studies that looked at how the amount of volunteering provided changed during the pandemic found that formal volunteering declined (Dederichs, 2022) and informal volunteering remained at a stable level (Cnaan et al., 2022). Mak and Fancourt (2021) report that older adults, people with more social support and people with higher education were doing more voluntary work during the pandemic than before. On the other hand, people living in urban areas were less likely to have increased their volunteering.

Given this short overview of volunteering during the pandemic, we now move on to literature that studied how the severity of a catastrophe affects prosocial behavior.

Prosocial behavior during catastrophes

There is evidence showing an increase in volunteering following disasters such as floods (Harris et al., 2017), hurricanes (Michel, 2007), terrorist attacks (Beyerlein & Sikkink, 2008), and foreclosures (Rotolo et al., 2015). This research also showed that the individuals who volunteered in response to these disasters displayed characteristics that are mostly similar to those of regular volunteers. However, Rotolo and Berg (2011) find that volunteers who volunteer for disaster relief tasks tend to be younger and less educated compared to volunteers who perform more general volunteering tasks. While there is evidence showing that disasters tend to lead to an increase in (disaster related) volunteering, the literature on how the severity of a disaster/crisis relates to the amount of volunteering during/after the crisis is very limited. To study the link between the severity of a crisis and volunteer turnout, Rotolo et al. (2015) used data from 120 U.S. metropolitan areas to assess the association between the foreclosure crisis and volunteering. They found that areas that experienced an increase in foreclosures also experienced an increase in volunteering rates. By using the variance in the severity of 57 disasters in Japan, Iizuka and Aldrich (2022) found that the number of deaths and missing persons as well as the size of the population affected by the disaster correlate most strongly with volunteer turnout.

Because the literature on the effect of the severity of crises/disasters on people's willingness to help in the form of volunteering is limited, we also make us of the donation literature to gain more insights into this relationship. While we acknowledge that the motives of volunteering and donating money differ in some ways, much of the literature deals with the concept of altruism as a motivation for both volunteering and donating money (Bekkers & Wiepking, 2011; J. Carpenter & Myers, 2010). This also holds true for the COVID-19 pandemic: Dury et al. (2022) found that altruism was the predominant motive for providing help during the first lockdown in Belgium, as 86.4% of the motives for providing help were linked to altruism.

Adena and Harke (2022) conducted an experiment to estimate the impact of COVID-19 severity on charitable donations using participants from England. They found that an additional 1% of cases resulted in an increase in donations by 2 to 11 pence. The authors suggest that increased awareness about COVID-19, due to higher local severity, was responsible for this effect. This interpretation was supported by a positive correlation between media coverage about the severity of COVID-19 and donations. Adding a COVID-19 reference to the donation request also increased donations. While Adena and Harke (2022) provide causal evidence that the severity of the COVID-19 pandemic positively affects donations, the study provides no insight into which psychological mechanism could be responsible for this effect. Because Adena and Harke (2022) attribute the effect to an increased awareness about COVID-19, empathy is a likely candidate. The increased awareness about COVID-19 and its negative effects on society could lead individuals to feel empathy with the victims of COVID-19, which would then lead to increased altruism reflected in the form of monetary donations or volunteering (a welldocumented link according to the empathy-altruism hypothesis (Batson et al., 1991)).

This interpretation is in line with the reasoning of Zheng et al. (2021), who studied the effect of the severity of collective threats on people's donation intention. Zheng et al. (2021) studied whether collective threats, such as the COVID-19 pandemic, positively affect people's intention to donate. They used environmental pollution and the COVID-19 pandemic to examine this effect and found that the severity of a collective threat has a positive effect on people's intention to donate to others facing the same threat. They found that this effect is serially mediated by people's other-focused attention and empathy. These results generalize nicely to our setting, as the volunteers in our sample also helped people that face the same threat as them and because Swiss people tended to see the COVID-19 pandemic as a collective threat (Albrecht et al., 2021).

While volunteering is a form of prosocial behavior and is likely motivated at least partly by similar motives (J. Carpenter & Myers, 2010), there is an important difference between donating and volunteering: the risk of the act of helping. A donation can be placed from the safety of one’s own home, while groceries cannot be delivered without getting into contact with others, which means subjecting oneself to the risk of getting infected with COVID-19. The risk of becoming infected is especially high if one is in an enclosed space with many other people (Noorimotlagh et al., 2021). This is the situation one encounters when going grocery shopping. The risk of becoming infected in situations such as this has been communicated and emphasized repeatedly by public health officials[7]. Thus, in the case of voluntary grocery deliveries, the positive effect of the severity of the COVID-19 pandemic on prosocial behavior (i.e., voluntary grocery deliveries) stands in conflict with the effect of one's own risk of becoming infected.

There is ample evidence showing that volunteers were afraid of becoming infected due to their volunteering activities during the COVID-19 pandemic (Cervera-Gasch et al., 2020; Lazarus et al., 2021). This very likely had a negative effect on volunteering. Ding et al. (2021) showed that fear of getting infected negatively affected intentions to perform voluntary work, but only if it included social contact. This effect was mediated by the state anxiety people experienced during the pandemic. A study by Rosychuk et al. (2008) also shows that risk perception had a negative effect on the decision to volunteer during an influenza pandemic. This negative effect is amplified by the evidence that shows that COVID-19 risk perception and prosociality are correlated (Dryhurst et al., 2020). Thus, given that volunteering in our case involved social contact and that our sample is by definition made up of prosocial individuals (i.e., volunteers), the fear of getting infected likely had a negative effect on volunteering in our setting. Therefore, while the reviewed literature suggests that rising case and death numbers should lead to more deliveries, this effect might be diminished by the rising risk of making deliveries,

[7] E.g., https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemienpandemien/aktuelle-ausbrueche-epidemien/novel-cov/so-schuetzen-wir-uns.html#490674923, accessed November 25, 2022

especially for prosocial individuals. Given that in earlier pandemics volunteers were willing to incur health risks for volunteering (Yonge et al., 2010), we don't think that this effect will completely offset the assumed positive effect of the severity of the pandemic.

Based on the reviewed literature that shows that the severity of disasters has a positive effect on prosocial behavior (i.e., volunteering), we hypothesize that the severity of the COVID19 pandemic, as measured by the case- and death numbers, had a positive effect on the amount of volunteering performed by individuals in our sample.

H1: There is a positive effect of weekly case and death numbers on the number of deliveries made per week.

A death from COVID-19 is more severe than simply contracting COVID-19 (i.e., a case). Given the literature that shows that the severity of disasters has a positive effect on prosocial behavior, we would therefore expect an increase in the death numbers to have a larger positive effect on volunteering than an equal increase in case numbers.

H2: The effect of the number of deaths on the number of deliveries made is larger than the effect of the number of cases on the number of deliveries made.

The reviewed literature also showed that the risk of getting infected with COVID-19 and the health impacts thereof negatively affect volunteering. We hypothesize that rising case- and death numbers will be perceived as a signal of the risk of volunteering and will therefore dampen the positive effect of the severity of the pandemic on volunteering. This hypothesis draws on the appraisal-tendency framework (Lerner & Keltner, 2000, 2001), which states that fear amplifies risk estimates. In the context of COVID-19, Harper et al. (2021) also found that fear of COVID19 and risk-perception were correlated. Thus, fear of not getting optimal treatment if one contracts COVID-19 likely amplifies the perceived risk of contracting COVID-19. As this fear is arguably higher when case- and death numbers are high and the health care system operates at or even over capacity, this amplification is stronger when case- and death numbers are high. The positive effect of threat severity on protective behavior (e.g., staying at home) is also formalized in theories like the health belief model (Janz & Becker, 1984) and the protection motivation theory (Rogers, 1975). These theories posit that the likelihood and magnitude of potential outcomes (e.g., likelihood of contracting COVID-19, magnitude of adverse health outcomes) affect protective behavior (Brewer et al., 2007). The positive association between protective health behavior and both the perceived likelihood of contracting a disease and its expected severity was demonstrated by meta-analyses (Brewer et al., 2007; C. J. Carpenter, 2010; Floyd et al., 2000). Transferred to our setting, these theories suggest that protective behavior (e.g., staying at home) increases as rising case- and death numbers lead to an increase in the likelihood and magnitude of adverse outcomes (i.e., potentially severe illness due to COVID-19). Based on these theories and evidence, we expect a concave effect of the case and death numbers on the number of deliveries made.

H3: The effect of the case and death numbers on the number of deliveries made is concave.

Of course, the case- and death numbers can only influence people's decision to volunteer if individuals are aware of those numbers. In a representative sample from Switzerland (collected in June 2020), half of the respondents knew the absolute number of COVID-19-related deaths, and a third of the respondents correctly stated the 7-day incidence rate (Albrecht et al., 2021). Combined with the broad covering of these numbers on the news, we are thus confident that this was the case.

In what follows, we present the data that allow us to test these hypotheses.

Methods

The ‘Amigos’ platform and data

Before describing the platform and data, we shortly provide some background about the Swiss context. According to the Swiss Volunteering Survey (Lamprecht et al., 2020), 39% of the population aged 15 and above engage in formal voluntary work through clubs and organizations, while 46% participate in informal voluntary work by serving as caretakers, assisting others, or supporting events and projects outside the scope of clubs and organizations. As Switzerland is a welfare state, individuals that suffered a loss of income due to the pandemic were supported by the government. Switzerland’s federal insurance programs that address sickness, unemployment, accidents, and old age work effectively and provide generous levels of benefits (Armingeon et al., 2021). Trust in others and especially in neighbors is high in Switzerland (Ortiz-Ospina & Roser, 2016).

The Amigos platform was launched as a cross-sector collaboration between Switzerland’s largest retailer (Migros) and Pro Senectute, Switzerland’s largest nonprofit for elderly people. The purpose of Pro Senectute is to maintain and enhance the well-being of the elderly in Switzerland. It does this by providing services for elderly both trough professionals and through volunteers. The Amigos platform allowed people who belonged to a risk group or who currently had to self-isolate to place grocery orders. These orders could then be delivered by volunteers who themselves signed up on the platform to do these deliveries. When volunteers signed up, they had to specify the geographical radius from which they wanted to be notified when an order was placed (Figure 1 left). Once the volunteers signed up, they saw a list of orders that were placed and are still open for delivery (Figure 1 right). Each item in the list shows the date and time window in which the order should be delivered and the zip code of the delivery address. It also shows how many items should be bought, how many shopping bags are (approximately) needed to carry the groceries, and the rating of the person who placed the order.

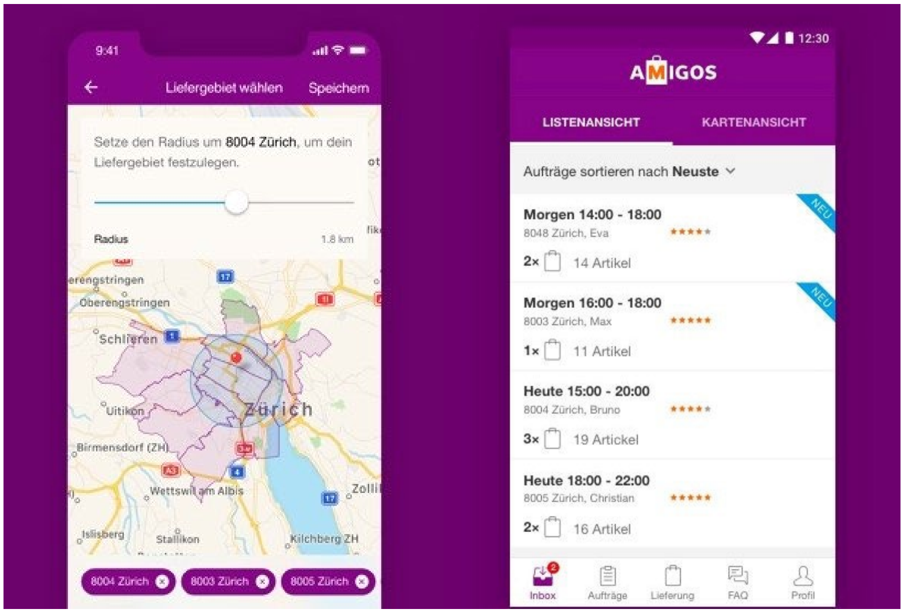


Figure 1 Left: When signing up, volunteers had to specify the radius from which they wanted to receive order notifications. Right: After signing up, volunteers see a list of open orders that they can choose to deliver.

Deliveries happened on a voluntary basis, but the person who placed the order could tip the delivery person via the app or in person. The volunteers were thus able to receive some form of remuneration, but the platform capped the maximum tip. The cap was first set to five Swiss francs. However, the operators of the platform noticed that almost all deliveries were tipped to the full five Swiss francs. Because of this ceiling effect, they decided to raise the cap to first seven and then nine Swiss francs. This remuneration is much below the remuneration paid by for-profit actors for similar services (e.g., Uber Eats) and much below the opportunity cost of volunteering (Wallrodt & Thieme, 2022). Therefore, by using the definition of volunteering by Cnaan et al. (1996), delivering groceries still qualifies as voluntary work. Migros provided us with an anonymized version of the data they collected.

Data

The data contain the date and status of each order (i.e., if the order has been successfully delivered and if yes by whom). For people who signed up to do deliveries, we have data on their age and sex. We also know the size of their delivery area (the zip codes they received notifications from when an order was placed). We augmented the Amigos data with COVID-19 case- and death number data provided by the Swiss Federal Office of Public Health. We first calculated the seven-day rolling average for both death and case numbers for each day before taking the mean of the resulting averages for each week. To control for the number of orders placed in a person's delivery area, we summed all orders that were placed in the person's delivery area in a given week. We use the number of deliveries made in a given week by a person as the dependent variable. This gives us an unbalanced panel data frame with a total of 26,960 individuals with up to 57 observations (i.e., weeks) per individual. A descriptive table of the data can be found in table 1 and will be discussed in more detail in the results section.

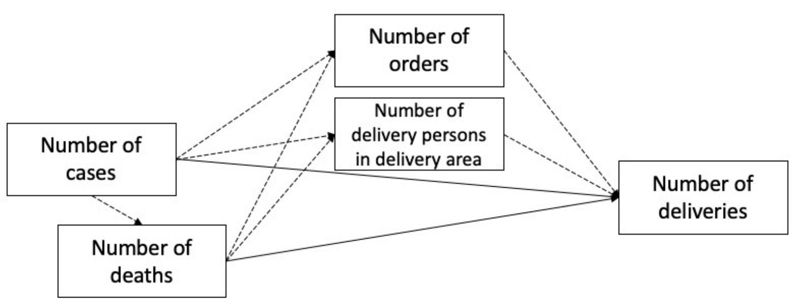


Figure 2 Depiction of the assumed causal relationship between the variables. Dotted lines represent indirect effects of the case- and death numbers on the number of deliveries made by a volunteer. Solid lines represent the direct effects that we isolate with the regression models.

Regression Models

To identify the effect of the case- and death numbers on the number of deliveries made by a volunteer, we need to control for the number of orders that were placed in a volunteer's delivery area. This is to control for the indirect paths of the case- and death numbers on the number of deliveries made that is mediated by the number of orders in a volunteer's delivery area. The same applies to the number of other delivery persons that are in a volunteer's delivery area (see Fig. 2). By doing this, we can estimate the direct effect of the case- and death numbers on the number of deliveries made by a volunteer. There is one caveat to our identification strategy: raising case numbers also means that more volunteers will become infected, rendering them unable to make deliveries. We would mistake the reduction in deliveries caused by this as a reduction in the willingness to do these deliveries. It might thus be that we slightly underestimate the effect of case numbers on deliveries if it is positive or overestimate the effect if it is negative. However, we believe that this bias is small, as the proportion of people who were infected in a given week was still rather small.

We use fixed effects regression models to estimate this direct effect by predicting the number of deliveries a volunteer made in a given week. The advantage of the fixed effects estimator is that it controls for unobserved heterogeneity on the level of volunteers that is constant over time (e.g., risk perception, trait empathy). As our dependent variable (DV) is the number of deliveries a volunteer made in a given week, we assume that the DV follows a Poisson distribution and thus use Poisson regression models. The Poisson model assumes that the mean and variance are the same, an assumption that is often violated. We nevertheless adhere to the Poisson model because it is more robust than models that incorporate an additional parameter to allow for over- or underdispersion (i.e., the negative binomial model) (Blackburn, 2015).

For the reasons mentioned above, all regression models control for the number of orders placed in the person's delivery area, the number of other delivery persons in the person's delivery area, and the interaction between these two. We run three model specifications: the first model uses case- and death numbers as independent variables, the second model only the death numbers and the third model only the case numbers. We run each of these models once with case- and death numbers that were aggregated over Switzerland and once with the local numbers. With local, we mean that we used the case- and death numbers of a volunteer's delivery area. These numbers were available on a cantonal level. If a volunteer's delivery areas spanned multiple cantons, we used the weighted average (weighted by the number of delivery areas in each canton) of the canton's case and death numbers. The models with the local case and death numbers also allow us to use time fixed effects in addition to individual fixed effects. These time fixed effects control for time specific shocks that affected all individuals (e.g., lockdowns). All regression models use Driscoll and Kraay (1998) standard errors, which are robust to heteroscedasticity, serial correlation, and cross-sectional dependence. The lag of the Driscoll and Kraay (1998) standard errors is based on Newey and West (1987). We use the coefficients from the linear effects of the case- and death numbers from the fixed effects models to test hypotheses 1 and 2 and the quadratic effects of these predictors to test hypothesis 3.

Results

Descriptive Results

Over the period the platform was in use, 26,960 volunteers signed up to deliver groceries. As seen in Figure 3 A), most of those signups happened during the early pandemic, i.e., while the first wave hit Switzerland. Figure 3 A) also shows that most of the volunteers who signed up did not end up delivering a single order. Only 7,569 volunteers who signed up also ended up making a delivery. These volunteers made a total of 72,379 deliveries. Figure 3 B) shows the weekly orders and deliveries. The figure reveals that most of the placed orders were also delivered. The largest relative mismatch between orders and deliveries could be observed in the summer, where both case- and death numbers were relatively low. Figures 3 C) and D) show the average weekly case and death numbers, respectively. The pattern of the weekly average number of deaths corresponds better to the number of deliveries than the pattern of the weekly average number of cases.

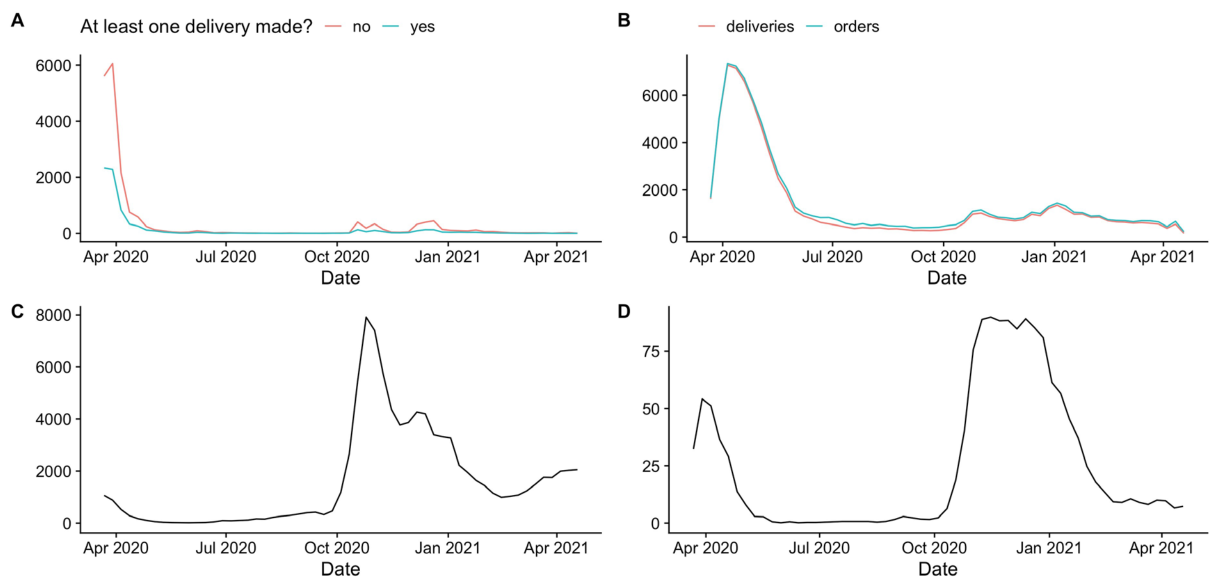


Figure 3 A) Number of volunteer signups per week. The color represents whether the volunteers who signed up ended up delivering at least one order or not. B) Number of orders and deliveries per week. C) Number of COVID-19 cases per week. D) Number of COVID-19 deaths per week.

Figure 4 A) shows the distribution of the total deliveries per person. Of those who made at least one delivery, most made no more than ten deliveries in total. However, the upper part of Figure 4 A) shows that some people ended up making hundreds of deliveries. Figure 4 B) visualizes the total number of deliveries made by people who made a given number of deliveries. This makes people who made many deliveries more visible because their count is now scaled by the total number of deliveries made by the given group of people who made the same number of total deliveries. Figure 4 B) shows that there are some people who ended up making many deliveries. To ensure that our results are not driven by these people, we exclude volunteers who ended up making more than 34.3 deliveries (mean + 2 SD). This removed a total of 384 people from our dataset (1.4%). While this is a small subgroup, it was responsible for almost half of all deliveries (35,310 deliveries, 48.7%). We report all results where these outliers are included in the appendix. Overall, the results are very similar for the models with and without outliers. A person who signed up on the platform ended up making 1.39 (Median = 0, SD = 3.99) deliveries on average. The subgroup of people who made at least one delivery on average ended up making 5.16 deliveries (Median = 2, SD = 6.29).

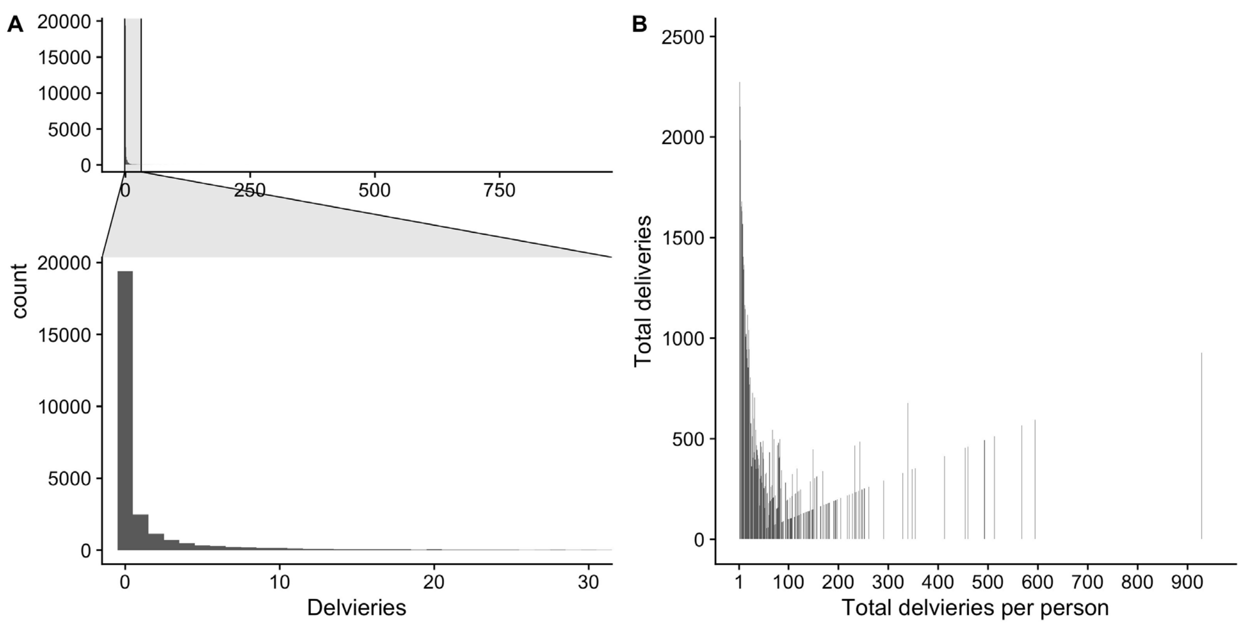


Figure 4 A) Histogram of the total number of deliveries made per person. Because some volunteers made a lot of deliveries, we zoomed into the distribution in the lower part of the figure as indicated by the grey shading in the upper part of the figure. B) Distribution of the total number of deliveries made by bins that represent volunteers who made a given number of total deliveries.

Table 1 lists summary statistics of the volunteer's individual characteristics, delivery characteristics and delivery area characteristics. The table is facetted by whether the volunteers ended up making at least one delivery or not. This allows us to see whether these two groups differ in substantial ways in any of the characteristics. The mean age of those who made at least one delivery was slightly lower than the mean age of those who did not make any deliveries. Although the difference is small, it is statistically significant. The distribution of sex was not significantly different between the group of people who made at least one delivery and those who did not. Those who ended up making at least one delivery on average received a tip of 5.44 Swiss Francs. Although most of the differences between the two groups are small in magnitude, they are still statistically significant due to the large N. Table 5 in the appendix reports the same results but grouped by outlier status. The nonoutlier group on average received a tip of 5.44 Swiss Francs, the outlier group received a slightly larger tip of 5.71 Swiss Francs on average. The two groups were also relatively similar with regard to the other characteristics reported in table 5. However, two notable differences emerged. First, the outlier group on average had a larger number of delivery areas (16.20) than nonoutliers (8.56). Second, volunteers from the outlier group on average signed up earlier (3.39 weeks after platform launch) than volunteers from the nonoutlier group (7.38 weeks after platform launch). Both factors likely contributed to the larger number of deliveries made by the volunteers in the outlier group.

Table 1 Descriptive Statistics

	At least one delivery made		
Variable	**False**, N = 19,391[1]	**True**, N = 7,185[1]	**p-value**[2]
Age			<0.001
Mean (IQR)	34.58 (25.79, 41.73)	33.57 (23.97, 41.60)	
Range	13.88, 88.53	15.51, 109.50	
Sex			0.017
F	13,016 (67.1%)	4,934 (68.7%)	
M	6,375 (32.9%)	2,251 (31.3%)	
Mean Tip per Week			
Mean (IQR)	NA (NA, NA)	5.44 (5.00, 5.73)	
Range	NA, NA	0.00, 9.00	
Number of delivery areas			<0.001
Mean (IQR)	7.71 (3.00, 11.00)	10.87 (5.00, 16.00)	
Range	1.00, 25.00	1.00, 25.00	
Mean number of orders in delivery area per week			<0.001
Mean (IQR)	2.24 (0.48, 3.62)	1.93 (0.48, 2.85)	
Range	0.00, 16.05	0.00, 16.05	
Mean number of pickers in delivery area per week			<0.001
Mean (IQR)	243.89 (92.85, 336.30)	210.01 (81.65, 265.92)	
Range	1.00, 950.58	2.66, 909.65	
Weeks between Signup and Launch of platform			0.035
Mean (IQR)	7.59 (0.00, 3.00)	6.81 (0.00, 3.00)	
Range	0.00, 56.00	0.00, 56.00	
Number of deliveries made			<0.001
Mean (IQR)	0.00 (0.00, 0.00)	5.16 (1.00, 6.00)	
Range	0.00, 0.00	1.00, 34.00	

[1] n (%)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test

Regression model results

Table 2 reports the estimated regression models. Looking at the coefficients of the case and death numbers, we see that the effects of the death numbers are more consistent across the models than the effects of the case numbers. The case numbers have a significant negative effect in the models with the aggregated case- and death numbers (models 1 & 3). But the significance of this effect vanishes once time fixed effects are introduced (model 4) or even turns significantly positive when the effect of the death numbers is not accounted for (model 6). A similar pattern holds for the effect of the squared case numbers that we used to check for the nonlinearity of the effect. The effect of the squared case numbers is positive (models 1 & 3) but again vanishes in the models with the local case- and death numbers (models 4 & 6). In contrast to the number of cases, the effects of the number of deaths are consistent across all models. Across all models (models 1, 2, 4 & 5), the number of deaths has a positive effect on the number of deliveries made per week. The sign of the squared death numbers is negative across all models (models 1, 2, 4 & 5), indicating a concave effect.

The fact that testing capacities increased over time (Federal Office of Public Health FOPH, 2022) could explain why the effect of the number of deaths is robust to the inclusion of time fixed effects, while the case numbers are not. In large part due to the increased testing capacity, case numbers were generally much higher in later stages of the pandemic than they were, e.g., in the first wave (Wu et al., 2020). Only the models with the time fixed effects control for this. We thus think that the negative effect of the case numbers on the number of deliveries made in the models with the aggregated case- and death numbers is an artifact of that. Because deaths are thus a more consistent measure of the severity of the pandemic, past research similar to ours opted to only use death numbers as a predictor (Fridman et al., 2022). However, for the sake of transparency, we opted to use both case- and death numbers.

The effect of a one-unit increase in the death numbers on the rate of deliveries depends on the value of the death numbers because of the exponential form of the Poisson model and because of the quadratic term. We therefore calculate the ratio of the expected rates of delivery for two consecutive case- or death numbers. The formula to do this is the following:

! $\dfrac{(Y|X = x + 1)}{E(Y|X = x)} = +$ ("!#""($%#&)),

where -& represents the coefficient of the linear effect and -$ reflects the coefficient of the quadratic effect. Notice that this ratio still depends on the value of x because of the quadratic term. A one-unit increase in the local death numbers when going from one to two local deaths (values close to the median) changes the rate of delivery by a factor of ~1.07 when using the coefficients from model (5) in Table 2. The effect for the same model but with lagged values (table 3) is slightly larger (~1.09). The values for the models where the case numbers are also included are very similar. Because of the quadratic effect of the death numbers, this factor gradually declines until an increase in the number of deaths leads to a decrease in the ratio of expected deliveries (see figure 5). This happens at around 11 local deaths for the model with the lagged deaths and at around 13 deaths for the models without lag.

Table 2 Regression Models

Dependent Variable:	Number of deliveries per week and person					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
number of delivery persons in delivery area	-0.0066***	-0.0084***	-0.0094***	-0.0008*	-0.0008*	-0.0009*
	(0.0009)	(0.0007)	(0.0010)	(0.0004)	(0.0004)	(0.0004)
number of orders in delivery area	0.1090***	0.2940***	0.3717***	0.0556***	0.0568***	0.0574***
	(0.0292)	(0.0355)	(0.0349)	(0.0068)	(0.0067)	(0.0066)
average cases per week/1000	-1.719***		-0.3782*	-0.6946		0.6741*
	(0.2325)		(0.1740)	(0.3905)		(0.3042)
average deaths per week/1000	83.18***	51.71*		85.92***	76.58***	
	(12.66)	(23.06)		(13.60)	(12.77)	
average cases per week^2/1000	0.0002***		4.9×10^{-5}*	0.0005		-0.0005
	(2.67×10^{-5})		(2.1×10^{-5})	(0.0003)		(0.0003)
average deaths per week^2/1000	-0.5007***	-0.5515*		-3.145***	-3.010***	
	(0.1473)	(0.2398)		(0.7474)	(0.7449)	
number of delivery persons in delivery area × number of orders in delivery area	-0.0002***	-0.0004***	-0.0005***	-0.0001***	-0.0001***	-0.0001***
	(4.11×10^{-5})	(5.84×10^{-5})	(6.95×10^{-5})	(1.32×10^{-5})	(1.31×10^{-5})	(1.24×10^{-5})
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Time				Yes	Yes	Yes
Case and Death numbers used	Aggregated	Aggregated	Aggregated	Local	Local	Local
Fit statistics						
Observations	360,635	360,635	360,635	360,635	360,635	360,635
Squared Correlation	0.27028	0.20299	0.19632	0.31831	0.31820	0.31781
Pseudo R^2	0.38207	0.34179	0.33109	0.42001	0.41998	0.41956
BIC	263,177.1	274,308.0	277,271.2	253,383.9	253,367.0	253,483.8

Driscoll-Kraay (L=2) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Looking at our hypotheses, we can only partially confirm hypothesis 1. Only the number of deaths, but not the number of cases, had a consistent positive effect on the number of deliveries made in all models. We can hence confirm hypothesis 2, namely, that the number of deaths has a larger effect on the number of deliveries made than the number of cases. Because only the effect of the number of deaths is concave, we can partially confirm hypothesis 3. The models discussed so far used the case- and death numbers of the week in which the deliveries were also made. However, it might be that the volunteers need some time to adapt their behavior to a change in case- and death numbers. We therefore fitted the same models with case- and death numbers lagged by one week. These models are presented in Table 3.

Table 3 Regression Models with lag of one week

Dependent Variable:	Number of deliveries per week and person					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
number of delivery persons in delivery area	-0.0158***	-0.0173***	-0.0133***	-0.0017*	-0.0017*	-0.0020*
	(0.0024)	(0.0016)	(0.0019)	(0.0007)	(0.0007)	(0.0008)
number of orders in delivery area	0.1364***	0.2565***	0.3970***	0.0458***	0.0462***	0.0470***
	(0.0249)	(0.0326)	(0.0403)	(0.0053)	(0.0053)	(0.0053)
average cases per week/1000	-1.152***		-0.2100	-0.4108		0.8577*
	(0.3274)		(0.1727)	(0.2861)		(0.3670)
average deaths per week/1000	87.57***	65.79**		106.0***	98.28***	
	(12.43)	(21.24)		(14.58)	(12.62)	
average cases per week^2/1000	0.0001**		2.83×10^{-5}	0.0005*		-0.0005
	(3.67×10^{-5})		(2.06×10^{-5})	(0.0002)		(0.0003)
average deaths per week^2/1000	-0.6426***	-0.6831**		-4.809***	-4.291***	
	(0.1668)	(0.2236)		(0.7695)	(0.7110)	
number of delivery persons in delivery area × number of orders in delivery area	-0.0003***	-0.0005***	-0.0006***	-0.0001***	-0.0001***	-0.0001***
	(5.66×10^{-5})	(6.92×10^{-5})	(9.09×10^{-5})	(1.17×10^{-5})	(1.19×10^{-5})	(1.22×10^{-5})
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Time				Yes	Yes	Yes
Case and Death numbers used	Aggregated	Aggregated	Aggregated	Local	Local	Local
Fit statistics						
Observations	305,777	305,777	305,777	305,777	305,777	305,777
Squared Correlation	0.28920	0.22760	0.20729	0.34336	0.34347	0.34224
Pseudo R^2	0.37489	0.34964	0.32616	0.41966	0.41963	0.41906
BIC	228,410.9	234,437.7	240,065.3	218,375.0	218,358.4	218,494.3

Driscoll-Kraay (L=2) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

The fit statistics are better for the models with the lagged case- and death numbers. This supports the conjecture that volunteers need some time to react to changes in case- and death numbers. Because the results are very similar to those reported in table 2 (which is to be expected since the case and death numbers are autocorrelated), we will not repeat them here and refer the reader to Table 3. This also means that the conclusion regarding our hypotheses remains unchanged.

Since gauging the nonlinearity of an effect from regression coefficients is difficult, we plotted the effect of the case- and death numbers in Figure 5. We do this only for the models with the local case and death numbers since we believe that these models are more robust. We also only plot the coefficients from models where both the linear and the nonlinear term were significant, this drops all case number coefficients. Plots for all regression models where both effects were significant are reported in the appendix for completeness. Figure 5 shows the concave effect of the number of deaths on the number of deliveries made that even turns slightly negative at the far end of the spectrum. However, as indicated by the boxplot on the top of the figure, the effect was positive for most of the time.

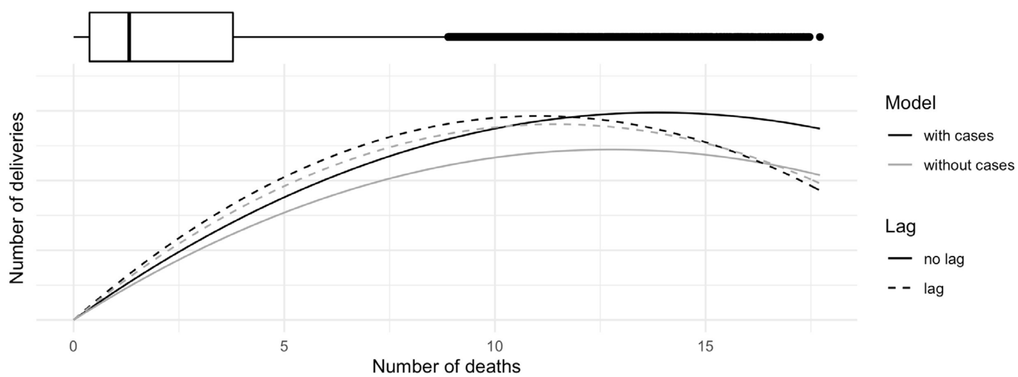


Figure 5 Plot of the combined linear and nonlinear effect of the local death numbers on the number of deliveries made as reported in Table 2 and 3. The box plot on the top of the figure visualizes the distribution of the weekly local death numbers. The y-axis has no numbered scale because the fixed effects models do not have an intercept.

Finally, looking at the coefficients of the control variables, we see that across all models, the number of other delivery persons in a volunteer's delivery area is negatively associated with the number of deliveries made per week. This is expected, as more delivery persons in a delivery area means that there are more people who can potentially claim an open delivery order. Unsurprisingly, the number of orders placed in a volunteer's delivery area is positively associated with the number of deliveries made per week across all models. The interaction between the last two reported variables is significantly negative across all models. Thus, the more other delivery persons there were in a volunteer's delivery area, the smaller the effect of an additional order. Overall, the signs of these control variables are all what one expects.

Discussion and Conclusion

In this work we tried to answer how the severity of the COVID-19 pandemic affected the provision of voluntary labor in the form of grocery food deliveries. By using fixed effects regression models, we found that the severity of the COVID-19 pandemic (i.e., the number of deaths) has a positive (but concave) effect on the amount of informal volunteer work provided by a given individual. The positive effect of the severity of the pandemic on the amount of volunteering is consistent with earlier, although cross-sectional, literature (e.g., Iizuka and Aldrich (2022)). However, past studies that investigated how the severity of the COVID-19 pandemic affected giving found a positive linear, and not concave, effect of COVID-19 severity on giving (Adena & Harke, 2022; Zheng et al., 2021). We assume that the rising risk of getting infected that is caused- and signaled by rising case- and death numbers is most likely to be responsible for the concavity of this effect. However, it is curious that we only find a concave effect for the death and not the case numbers, since cases and not deaths cause infections. It could be that higher death numbers signaled a higher risk of negative health events when contracting COVID-19. This would be in line with the introduced literature that shows that fear amplifies risk perception (Lerner & Keltner, 2000, 2001). Evidence from the Health Belief Model could also explain why only the death numbers had a concave effect. A meta-analysis found that that the perceived severity had a larger effect on protective health behaviors than the perceived susceptibility (C. J. Carpenter, 2010). It could be that rising death numbers and news about hospitals reaching capacity made the severity of contracting COVID-19 more salient. As rising case numbers likely only have an effect on perceived susceptibility, the positive effect of rising death numbers on protective health behaviors would be higher than the one of rising case numbers.

A reason why we find different results for case and death numbers might be that people were more motivated to volunteer at the beginning of the pandemic. Indeed, most deliveries were made in the early parts of the pandemic. During that time, the death-to-case ratio was considerably higher than in the second period with many deliveries (winter 2020/2021). The models with the time fixed effects control for this, and the fact that the negative effect of the case numbers vanishes in these models is in line with this conjecture.

There are also other mechanisms that could lead to the observed concave effect, e.g., compassion fade (Butts et al., 2019), and our data again do not allow us to rule out one or the other. However, the widespread media coverage about the risks of COVID-19 (Silini, 2020), the correlation between prosociality and risk perception (Dryhurst et al., 2020), and the fact that risk perception negatively influenced the decision to volunteer in past pandemics (Rosychuk et al., 2008) all point to risk perception being the most plausible explanation. The fact that studies that looked at how the severity of the pandemic affected giving did not find a nonlinear effect (e.g., Adena & Harke, 2021) also speaks for risk perception, as compassion fade should affect both giving and volunteering in the same way. This interpretation is also in line with the finding of Mak and Fancourt (2021) that people with a diagnosed disability or illness had lower odds of volunteering in neighborhood support during the COVID-19 pandemic (preexisting conditions are a risk factor for severe COVID-19 outcomes, Jordan et al., 2020). Finally, risk perception has been found to play a role in an earlier survey study conducted on a sample of the population studied here. Trautwein et al. (2020) studied under which conditions volunteers were satisfied with their COVID-19 volunteering mediated by platforms such as Amigos. They found that the perceived susceptibility to a COVID-19 infection moderated the relationship between the evaluation of the crisis-policy measures implemented by the platform and satisfaction with the volunteering experience.

Regarding the implications for practice, our study provides evidence that online platforms such as Amigos are an effective and efficient way of matching spontaneous volunteers with people who need help, as more than 90% of the orders that were placed were also delivered. Thus, such platforms prove to be an effective way to address some of the challenges associated with informal volunteering in emergencies and disasters (Daddoust et al., 2021; Whittaker et al., 2015). For example, such platforms might be a viable solution to deal with the problem of an oversupply of volunteers (Simsa et al., 2019). By taking care of the matching before volunteers arrive on site, oversupply is by design not possible. This reduces disappointment on the side of the volunteers and at the same time strengthens confidence in asking for help by beneficiaries. While matching in our case was quite easy because the type of activity was quite simple, it is easy to imagine the use of such platforms in other cases. For example, after a flooding, a platform might match people whose homes have been demolished with people who are willing to help with cleanup/rebuilding. The people affected could list the activities to be performed, the expected duration and the number of volunteers needed, and the platform could facilitate matching based on these criteria. Schmidt and Albert (2022) provide proof of the feasibility of such a paradigm where volunteers self-assign to tasks from an ordered list of recommendations. Betke (2018) lays out in detail how such a platform might work and how it could be successfully integrated into existing structures. In a very recent event, Swiss citizens signed up on online platforms that matched Ukrainian refugees with people who were willing to temporarily host these refugees in their homes (Brelie, 2022). By providing low-barrier access to helping opportunities, platforms like Amigos could also be an effective tool to counteract the decline in formal volunteering seen in the early stages of the pandemic (Dederichs, 2022). While it might have been true in the past that spontaneous volunteers cannot be actively “harvested” (KoolenMaas et al., 2022), our results show that app-mediated active harvesting (i.e., recruiting) of spontaneous volunteers can be done with great success. Such apps can thus be a valuable tool to manage the different types of volunteer resources (Koolen-Maas et al., 2022).

To mitigate the negative effect of the risk of volunteering, such platforms should provide volunteers with information on how to minimize the risk caused by volunteering (e.g., wearing masks). Indeed, in an earlier study conducted on a sample of individuals that used the Amigos platform to volunteer, Trautwein et al. (2021) found that the evaluation of the platforms' crisispolicy measures (e.g., supply of health information) had a positive impact on COVID-19 volunteering satisfaction. This is in line with the findings of Rosychuk et al. (2008).

Given the nature of our dataset covering only one task and one country might limit the generalizability of our findings. However, delivering groceries was the most frequent informal volunteering task during the pandemic in other countries (e.g., Mao et al., 2021), and there is no reason to expect that our proposed underlying mechanism that drives our results should radically differ across countries. As our research concerns behavior during a pandemic/crisis, it is not clear whether or how these findings generalize to normal times. Generally, our study supports existing studies that emphasize the use of technology to improve volunteer matching in normal times (Chui & Chan, 2019). The fact that the demographic profile of COVID-19 volunteers resembled that of people who volunteer during normal times (Mak & Fancourt, 2021) promises at least cautious transferability. This also alleviates concerns regarding a selection bias of our sample. However, we acknowledge that there could still be selection bias, as we do not know how the group of people who signed up on the platform compares to the group of people who did not. It is possible that the group who signed up could have been more prosocial and/or less risk averse than the group of people who did not sign up. While not ruling out selection, the fact that the platform was designed and advertised by two of the most well-known organizations in Switzerland should have ensured that the platform was known by many people. Last, we did not consider the amount of media coverage of the service, which could have influenced the number of volunteers that signed up on the platform and the number of deliveries made by these volunteers.

References

Adena, M., & Harke, J. (2022). Covid-19 and pro-sociality: How do donors respond to local pandemic severity, increased salience, and media coverage? *Experimental Economics*, *25*(3), 824–844. https://doi.org/10.1007/s10683-022-09753-y

Albrecht, R., Jarecki, J. B., Meier, D. S., & Rieskamp, J. (2021). Risk preferences and risk perception affect the acceptance of digital contact tracing. *Humanities and Social Sciences Communications*, *8*(1). https://doi.org/10.1057/s41599-021-00856-0

Armingeon, K., Sager, F., & Zohlnhöfer, R. (2021). *Switzerland Report : Sustainable Governance in the Context of the COVID-19 Crisis*. Gütersloh. https://doi.org/10.11586/2021107

Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *61*(3), 413–426. https://doi.org/10.1037/0022-3514.61.3.413

Bekkers, R., & Wiepking, P. (2011). A Literature Review of Empirical Studies of Philanthropy. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973. https://doi.org/10.1177/0899764010380927

Betke, H. (2018). A Volunteer Coordination System Approach for Crisis Committees. In *ISCRAM*.

Beyerlein, K., & Sikkink, D. (2008). Sorrow and Solidarity: Why Americans Volunteered for 9/11 Relief Efforts. *Social Problems*, *55*(2), 190–215. https://doi.org/10.1525/sp.2008.55.2.190

Blackburn, M. L. (2015). The Relative Performance of Poisson and Negative Binomial Regression Estimators. *Oxford Bulletin of Economics and Statistics*, *77*(4), 605–616.

https://doi.org/10.1111/obes.12074

Brelie, J. von der. (2022). *Resources and support for those affected by the war in Ukraine (No. CERN-HR-Note-2022-008).*

Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, *26*, 136–145. https://doi.org/10.1037/0278-6133.26.2.136

Brudney, J. L., Handy, F., & Meijs, L. C. P. M. (2019). Introduction to the Special Issue: Exploring the Dimensions of Volunteering. *Nonprofit and Voluntary Sector Quarterly*, *48*(2_suppl), 5S-11S. https://doi.org/10.1177/0899764018808981

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16–33. https://doi.org/10.1016/j.obhdp.2018.12.006

Carlsen, H. B., Toubøl, J., & Brincker, B. (2020). On solidarity and volunteering during the COVID-19 crisis in Denmark: the impact of social networks and social media groups on the distribution of support. *European Societies*, 1–19. https://doi.org/10.1080/14616696.2020.1818270

Carpenter, C. J. (2010). A Meta-Analysis of the Effectiveness of Health Belief Model Variables in Predicting Behavior. *Health Communication*, *25*(8), 661–669. https://doi.org/10.1080/10410236.2010.521906

Carpenter, J., & Myers, C. K. (2010). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics*, *94*(11-12), 911–920. https://doi.org/10.1016/j.jpubeco.2010.07.007

Cervera-Gasch, Á., González-Chordá, V. M., & Mena-Tudela, D. (2020). COVID-19: Are Spanish medicine and nursing students prepared? *Nurse Education Today*, *92*, 104473.

Chui, C. H.-K., & Chan, C. H. (2019). The role of technology in reconfiguring volunteer management in nonprofits in Hong Kong: Benefits and discontents. *Nonprofit Management and Leadership*, *30*(1), 89–111. https://doi.org/10.1002/nml.21369

Cnaan, R. A., Handy, F., Marrese, T., Choi, D., & Ferris, A. (2022). *GENEROSITY TRENDS AND IMPACTS: Before and During the COVID-19 Pandemic in the USA.* Penn Social Policy & Practice. https://www.sp2.upenn.edu/wp-content/uploads/2014/07/Generosity-Trends-and-Impacts.pdf

Cnaan, R. A., Handy, F., & Wadsworth, M. (1996). Defining Who is a Volunteer: Conceptual and Empirical Considerations. *Nonprofit and Voluntary Sector Quarterly*, *25*(3), 364–383. https://doi.org/10.1177/0899764096253006

Daddoust, L., Asgary, A., McBey, K. J., Elliott, S., & Normand, A. (2021). Spontaneous volunteer coordination during disasters and emergencies: Opportunities, challenges, and risks. *International Journal of Disaster Risk Reduction*, *65*, 102546. https://doi.org/10.1016/j.ijdrr.2021.102546

Dederichs, K. (2022). Volunteering in the United Kingdom During the COVID-19 Pandemic: Who Started and Who Quit? *Nonprofit and Voluntary Sector Quarterly*, 089976402211228. https://doi.org/10.1177/08997640221122814

Ding, Y., Ji, T., & Guo, Y. (2021). Helping While Social Distancing: Pathogen Avoidance Motives Influence People's Helping Intentions during the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*, *18*(22). https://doi.org/10.3390/ijerph182212113

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, *80*(4), 549–560.

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., Spiegelhalter, D., & van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, *23*(7-8), 994–1006. https://doi.org/10.1080/13669877.2020.1758193

Dury, S., Brosens, D., Pan, H., Principi, A., Smetcoren, A.-S., Perek-Białas, J., & Donder, L. de (2022). Helping Behavior of Older Adults during the Early COVID-19 Lockdown in Belgium. *Research on Aging*, 1640275221105231. https://doi.org/10.1177/01640275221105231

Federal Office of Public Health FOPH. (2022, November 1). *Tests and share of positive tests*. Covid-19 Switzerland: Coronavirus: Dashboard. https://www.covid19.admin.ch/en/epidemiologic/test?time=total&epiRelDev=abs Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A Meta-Analysis of Research on Protection Motivation Theory. *Journal of Applied Social Psychology*, *30*(2), 407–429. https://doi.org/10.1111/j.1559-1816.2000.tb02323.x

Fridman, A., Gershon, R., & Gneezy, A. (2022). Increased generosity under COVID-19 threat. *Scientific Reports*, *12*(1), 4886. https://doi.org/10.1038/s41598-022-08748-2

Harper, C. A., Satchell, L. P., Fido, D., & Latzman, R. D. (2021). Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. *International Journal of Mental Health and Addiction*, *19*(5), 1875–1888. https://doi.org/10.1007/s11469-020-00281-5

Harris, M., Shaw, D., Scully, J., Smith, C. M., & Hieke, G. (2017). The Involvement/Exclusion Paradox of Spontaneous Volunteering. *Nonprofit and Voluntary Sector Quarterly*, *46*(2), 352–371. https://doi.org/10.1177/0899764016654222

Iizuka, A., & Aldrich, D. P. (2022). Attracting altruists: Explaining volunteer turnout during natural hazard events in Japan. *Disasters*, *46*(2), 526–544. https://doi.org/10.1111/disa.12477

Janz, N. K., & Becker, M. H. (1984). The Health Belief Model: A decade later. *Health Education Quarterly*, *11*, 1–47. https://doi.org/10.1177/109019818401100101

Jones, M., Beardmore, A., Biddle, M., Gibson, A., Ismail, S. U., McClean, S., & White, J. (2020). Apart but not Alone? A cross-sectional study of neighbour support in a major UK urban area during the COVID-19 lockdown. *Emerald Open Research*, *2*, 37. https://doi.org/10.35241/emeraldopenres.13731.1

Jordan, R. E., Adab, P., & Cheng, K. K. (2020). Covid-19: Risk factors for severe disease and death. *BMJ (Clinical Research Ed.)*, *368*, m1198. https://doi.org/10.1136/bmj.m1198

Koolen-Maas, S. A., Meijs, L. C. P., van Overbeeke, P. S. M., & Brudney, J. L. (2022). Rethinking Volunteering as a Natural Resource: A Conceptual Typology. *Nonprofit and Voluntary Sector Quarterly*, 089976402211279. https://doi.org/10.1177/08997640221127947

Lamprecht, M., Fischer, A., & Stamm, H. (2020). *Freiwilligen-Monitor Schweiz*. Seismo.

Lazarus, G., Findyartini, A., Putera, A. M., Gamalliel, N., Nugraha, D., Adli, I., Phowira, J., Azzahra, L., Ariffandi, B., & Widyahening, I. S. (2021). Willingness to volunteer and readiness to practice of undergraduate medical students during the COVID-19 pandemic: a cross-sectional survey in Indonesia. *BMC Medical Education*, *21*(1), 138. https://doi.org/10.1186/s12909-021-02576-0

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, *14*(4), 473–493. https://doi.org/10.1080/026999300402763

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*, 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Mak, H. W., & Fancourt, D. (2021). Predictors of engaging in voluntary work during the COVID-19 pandemic: Analyses of data from 31,890 adults in the UK. *Perspectives in Public Health*, 1757913921994146. https://doi.org/10.1177/1757913921994146

Mao, G., Fernandes-Jesus, M., Ntontis, E., & Drury, J. (2021). What have we learned about COVID-19 volunteering in the UK? A rapid review of the literature. *BMC Public Health*, *21*(1), 1470. https://doi.org/10.1186/s12889-021-11390-8

Meyersohn, N. (2020). *Coronavirus will change the grocery industry forever.* CNN Business. https://edition.cnn.com/2020/03/19/business/grocery-shopping-onlinecoronavirus/index.html

Michel, L. M. (2007). Personal responsibility and volunteering after a natural disaster: The case of Hurricane Katrina. *Sociological Spectrum*, *27*(6), 633–652. https://doi.org/10.1080/02732170701533855

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708. http://www.jstor.org/stable/1913610

NHS. (2020). *'Your NHS Needs You' – NHS call for volunteer army.* NHS. https://www.england.nhs.uk/2020/03/your-nhs-needs-you-nhs-call-for-volunteer-army/

Noorimotlagh, Z., Jaafarzadeh, N., Martínez, S. S., & Mirzaee, S. A. (2021). A systematic review of possible airborne transmission of the COVID-19 virus (SARS-CoV-2) in the indoor air environment. *Environmental Research*, *193*, 110612. https://doi.org/10.1016/j.envres.2020.110612

Ortiz-Ospina, E., & Roser, M. (2016). *Trust.* Our World in Data. https://ourworldindata.org/trust

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change1. *The Journal of Psychology*, *91*(1), 93–114. https://doi.org/10.1080/00223980.1975.9915803

Rosychuk, R. J., Bailey, T., Haines, C., Lake, R [Robert], Herman, B., Yonge, O., & Marrie, T. J. (2008). Willingness to volunteer during an influenza pandemic: Perspectives from students and staff at a large Canadian university. *Influenza and Other Respiratory Viruses*, *2*(2), 71–79. https://doi.org/10.1111/j.1750-2659.2008.00042.x

Rotolo, T., & Berg, J. A. (2011). In times of need: An examination of emergency preparedness and disaster relief service volunteers. *Nonprofit and Voluntary Sector Quarterly*, *40*(4), 740-750. https://doi.org/ 10.1177/0899764010369179

Rotolo, T., Wilson, J., & Dietz, N. (2015). Volunteering in the United States in the Aftermath of the Foreclosure Crisis. *Nonprofit and Voluntary Sector Quarterly*, *44*(5), 924–944. https://doi.org/10.1177/0899764014546669

Schmidt, A. P., & Albert, L. A. (2022). Task recommendations for self-assigning spontaneous volunteers. *Computers & Industrial Engineering*, *163*, 107798. https://doi.org/10.1016/j.cie.2021.107798

Silini, A. (2020, June 12). Switzerland: Coronavirus and the media. European Journalism Observatory. https://en.ejo.ch/ethics-quality/switzerland-coronavirus-and-the-media

Simsa, R., Rameder, P., Aghamanoukjan, A., & Totter, M. (2019). Spontaneous Volunteering in Social Crises: Self-Organization and Coordination. *Nonprofit and Voluntary Sector Quarterly*, *48*(2), 103–122. https://doi.org/10.1177/0899764018785472

Trautwein, S., Liberatore, F., Lindenmeier, J., & Schnurbein, G. von (2020). Satisfaction With Informal Volunteering During the COVID-19 Crisis: An Empirical Study Considering a Swiss Online Volunteering Platform. *Nonprofit and Voluntary Sector Quarterly*, *49*(6), 1142–1151. https://doi.org/10.1177/0899764020964595

UNRIC Brussels. *COVID-19 drives global surge in volunteering.* United Nations. https://www.un.org/en/coronavirus/covid-19-drives-global-surge-volunteering

Wallrodt, S., & Thieme, L. (2022). The Opportunity Costs of Volunteering: Evidence From Germany. *Nonprofit and Voluntary Sector Quarterly*, 089976402211295. https://doi.org/10.1177/08997640221129579

Whittaker, J., McLennan, B., & Handmer, J. (2015). A review of informal volunteerism in emergencies and disasters: Definition, opportunities and challenges. *International Journal of Disaster Risk Reduction*, *13*, 358–368. https://doi.org/10.1016/j.ijdrr.2015.07.010

Wilson, J. (2012). Volunteerism Research. *Nonprofit and Voluntary Sector Quarterly*, *41*(2), 176–212. https://doi.org/10.1177/0899764011434558

Wu, S. L., Mertens, A. N., Crider, Y. S., Nguyen, A., Pokpongkiat, N. N., Djajadi, S., Seth, A., Hsiang, M. S., Colford, J. M., Reingold, A., Arnold, B. F., Hubbard, A., & BenjaminChung, J. (2020). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, *11*(1), 4507. https://doi.org/10.1038/s41467-020-18272-4

Yonge, O., Rosychuk, R. J., Bailey, T. M., Lake, R [Rob], & Marrie, T. J. (2010). Willingness of University Nursing Students to Volunteer During a Pandemic. *Public Health Nursing*, *27*(2), 174–180. https://doi.org/10.1111/j.1525-1446.2010.00839.x

Zaki, J. (2020). Catastrophe Compassion: Understanding and Extending Prosociality Under Crisis. *Trends in Cognitive Sciences*, *24*(8), 587–589. https://doi.org/10.1016/j.tics.2020.05.006

Zheng, C., Liu, N., Luo, C., & Wang, L. (2021). Effects of the severity of collective threats on people's donation intention. *Psychology & Marketing*, *38*(9), 1426–1439. https://doi.org/10.1002/mar.21536

Appendix

Table 4 Descriptive statistics table with outliers

Variable	At least one delivery made: **False**, N = 19,391[1]	**True**, N = 7,569[1]	p-value[2]
Age			<0.001
Mean (IQR)	34.58 (25.79, 41.73)	33.80 (24.15, 41.91)	
Range	13.88, 88.53	15.51, 109.50	
Sex			0.16
F	13,016 (67.1%)	5,149 (68.0%)	
M	6,375 (32.9%)	2,420 (32.0%)	
Mean Tip per Week			
Mean (IQR)	NA (NA, NA)	5.45 (5.00, 5.83)	
Range	NA, NA	0.00, 9.00	
Number of delivery areas			<0.001
Mean (IQR)	7.71 (3.00, 11.00)	11.14 (5.00, 17.00)	
Range	1.00, 25.00	1.00, 25.00	
Mean number of orders in delivery area per week			<0.001
Mean (IQR)	2.24 (0.48, 3.62)	1.95 (0.49, 2.89)	
Range	0.00, 16.05	0.00, 16.05	
Mean number of pickers in delivery area per week			<0.001
Mean (IQR)	243.89 (92.85, 336.30)	210.89 (82.64, 267.66)	
Range	1.00, 950.58	2.66, 909.65	
Weeks between Signup and Launch of platform			<0.001
Mean (IQR)	7.59 (0.00, 3.00)	6.63 (0.00, 3.00)	
Range	0.00, 56.00	0.00, 56.00	
Number of deliveries made			<0.001
Mean (IQR)	0.00 (0.00, 0.00)	9.56 (1.00, 8.00)	
Range	0.00, 0.00	1.00, 928.00	

[1] n (%)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test

Table 5 Comparing outliers with nonoutliers

Variable	At least one delivery made: False, N = 19,391[1]	At least one delivery made: True, N = 7,185[1]	p-value[2]
Age			<0.001
Mean (IQR)	34.58 (25.79, 41.73)	33.57 (23.97, 41.60)	
Range	13.88, 88.53	15.51, 109.50	
Sex			0.017
F	13,016 (67.1%)	4,934 (68.7%)	
M	6,375 (32.9%)	2,251 (31.3%)	
Mean Tip per Week			
Mean (IQR)	NA (NA, NA)	5.44 (5.00, 5.73)	
Range	NA, NA	0.00, 9.00	
Number of delivery areas			<0.001
Mean (IQR)	7.71 (3.00, 11.00)	10.87 (5.00, 16.00)	
Range	1.00, 25.00	1.00, 25.00	
Mean number of orders in delivery area per week			<0.001
Mean (IQR)	2.24 (0.48, 3.62)	1.93 (0.48, 2.85)	
Range	0.00, 16.05	0.00, 16.05	
Mean number of pickers in delivery area per week			<0.001
Mean (IQR)	243.89 (92.85, 336.30)	210.01 (81.65, 265.92)	
Range	1.00, 950.58	2.66, 909.65	
Weeks between Signup and Launch of platform			0.035
Mean (IQR)	7.59 (0.00, 3.00)	6.81 (0.00, 3.00)	
Range	0.00, 56.00	0.00, 56.00	
Number of deliveries made			<0.001
Mean (IQR)	0.00 (0.00, 0.00)	5.16 (1.00, 6.00)	
Range	0.00, 0.00	1.00, 34.00	

[1] n (%)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test

Table 6 *Regression table with outliers*

Dependent Variable:	Number of deliveries per week and person					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
number of delivery persons in delivery area	-0.0009	-0.0039***	-0.0038***	0.0026***	0.0026***	0.0025***
	(0.0013)	(0.0009)	(0.0006)	(0.0005)	(0.0005)	(0.0004)
number of orders in delivery area	0.1168***	0.2646***	0.3015***	0.0439***	0.0445***	0.0447***
	(0.0317)	(0.0279)	(0.0214)	(0.0085)	(0.0087)	(0.0081)
average cases per week/1000	-1.344***		-0.3422*	-0.3284		0.3293
	(0.2244)		(0.1563)	(0.3368)		(0.2547)
average deaths per week/1000	62.14***	29.62		47.58***	40.61***	
	(12.03)	(19.09)		(9.162)	(7.982)	
average cases per week^2/1000	0.0001***		4.36×10^{-5}*	0.0005		-2.12×10^{-5}
	(2.5×10^{-5})		(1.83×10^{-5})	(0.0003)		(0.0002)
average deaths per week^2/1000	-0.3743**	-0.3270		-2.056***	-1.460**	
	(0.1284)	(0.2002)		(0.5047)	(0.4594)	
number of delivery persons in delivery area × number of orders in delivery area	-0.0002***	-0.0004***	-0.0005***	-0.0001***	-0.0001***	-0.0001***
	(3.64×10^{-5})	(3.45×10^{-5})	(4.81×10^{-5})	(1.85×10^{-5})	(1.88×10^{-5})	(1.75×10^{-5})
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Time				Yes	Yes	Yes
Case and Death numbers used	Aggregated	Aggregated	Aggregated	Local	Local	Local
Fit statistics						
Observations	381,221	381,221	381,221	381,221	381,221	381,221
Squared Correlation	0.31740	0.27243	0.30317	0.39840	0.39837	0.39801
Pseudo R^2	0.49319	0.46481	0.46446	0.52718	0.52714	0.52706
BIC	356,148.9	370,612.2	370,793.7	339,511.1	339,507.2	339,549.3

Driscoll-Kraay (L=2) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Table 7 Regression table with outliers and lag of one week

Dependent Variable:	Number of deliveries per week and person					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
number of delivery persons in delivery area	-0.0055*	-0.0081***	-0.0041**	0.0032***	0.0033***	0.0031***
	(0.0025)	(0.0017)	(0.0015)	(0.0006)	(0.0006)	(0.0005)
number of orders in delivery area	0.1234***	0.2304***	0.3145***	0.0331***	0.0335***	0.0338***
	(0.0296)	(0.0292)	(0.0248)	(0.0070)	(0.0071)	(0.0068)
average cases per week/1000	-1.042***		-0.2454	-0.3568		0.3379
	(0.2859)		(0.1500)	(0.3711)		(0.3077)
average deaths per week/1000	68.85***	43.85*		50.11***	42.33***	
	(12.24)	(18.95)		(11.84)	(9.682)	
average cases per week^2/1000	0.0001***		3.19×10^{-5}	0.0005		-2.13×10^{-5}
	(3.15×10^{-5})		(1.78×10^{-5})	(0.0003)		(0.0003)
average deaths per week^2/1000	-0.4910***	-0.4677*		-2.086***	-1.467***	
	(0.1429)	(0.2009)		(0.5063)	(0.4404)	
number of delivery persons in delivery area × number of orders in delivery area	-0.0003***	-0.0004***	-0.0005***	-9.54×10^{-5}***	-9.57×10^{-5}***	-9.82×10^{-5}***
	(4.8×10^{-5})	(4.31×10^{-5})	(6.12×10^{-5})	(1.85×10^{-5})	(1.86×10^{-5})	(1.82×10^{-5})
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Time				Yes	Yes	Yes
Case and Death numbers used	Aggregated	Aggregated	Aggregated	Local	Local	Local
Fit statistics						
Observations	325,979	325,979	325,979	325,979	325,979	325,979
Squared Correlation	0.36105	0.29509	0.31640	0.42295	0.42297	0.42245
Pseudo R^2	0.50064	0.47875	0.46980	0.53680	0.53675	0.53665
BIC	315,247.2	325,364.5	329,508.8	299,188.5	299,185.2	299,233.3

Driscoll-Kraay (L=2) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

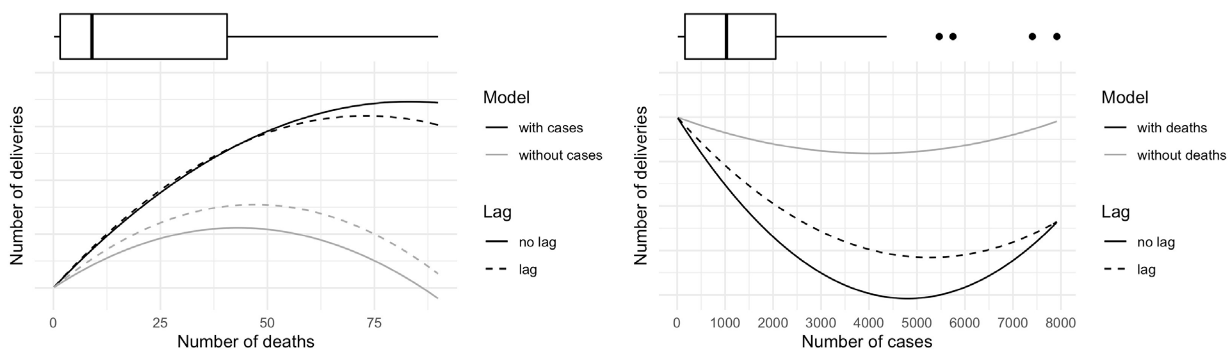


Figure 6 Plot of the combined linear and nonlinear effect of the aggregated death and case numbers on the number of deliveries made as reported in Table 2 and 3. Only models where both the linear and nonlinear effect is significant are shown. The box plot on the top of the figure visualizes the distribution of the weekly death and case numbers. The y-axis has no numbered scale because the fixed effects models do not have an intercept.