HIAS-E-139

Charity Fraud: An Experimental Study of the Moral Hazard Problem in the Charity Market

Hiroki Kato[(a)], Youngrok Kim [(b)]

(a) Hitotsubashi University
(b) University of Hyogo

April 2, 2024

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8668 http://hias.hit-u.ac.jp/

HIAS discussion papers can be downloaded without charge from:
https://hdl.handle.net/10086/27202
https://ideas.repec.org/s/hit/hiasdp.html

Abstract

Donors entrust their resources to charitable organizations, which use these to carry out activities that contribute to society. Thus, the donation market carries the potential for moral hazard on the part of charitable organizations. However, empirical insights into the negative consequences of this problem are scarce. This study creates a unique game by incorporating elements of principal-agent relationships into the dictator game, in order to quantitatively examine the consequences of moral hazard in the donation market through laboratory experiments. The results reveal that moral-hazard environments hinder the average donation amount from donors and, furthermore, decrease recipients' average payoffs. To address this negative outcome, non-binding promises are introduced to inform donors beforehand of the actions to be taken by charitable organizations. Although these promises succeed in offsetting the decrease in donation amounts, they do not sufficiently improve recipient welfare.

Keywords: Charity market; Dictator game; Fraud; Hidden action; Promise

JEL codes: C71; C92; D64; D82

1 Introduction

Despite charitable organizations prioritizing transparency, the charity market harbors a potential moral hazard problem. Sharing asymmetric characteristics with corporations (shareholders-managers), donors may be unaware of the consequences of their donations to charitable organizations. Consequently, the donation market faces a potential, classical principal-agent problem. Although this point is by no means a new perspective (e.g., Rose-Ackerman, 1996), however, a scarcity of insights exists into this issue's negative consequences.

The key difference with classic principal-agent relationships is that charitable organizations do not pursue donors' economic interests. Shareholders entrust some of their funds to managers, who utilize the funds to generate profits and return them to the shareholders (Jensen and Meckling, 1976). Conversely, donors entrust some of their funds to charitable organizations, which utilize these funds to benefit society.

Given this difference, principals (shareholders or donors) invest with different motives. Shareholders invest with selfish motives, whereas donors invest with altruistic motives contributing to society. A similarity is that agents, exploiting the hidden action environment where principals cannot observe or verify agent actions, can pursue private benefits. Some charitable organizations may aim not only for social contribution but also for private gain.

In such a structure, fraud can occur even in nonprofit organizations where collected funds are misused. Archambeault et al. (2015) found that 17% of scandals involving nonprofit organizations in the United States fall into such categories. In severe instances, charity fraud—such as masquerading as fundraising for disaster relief or humanitarian aid—frequently occur. Hence, charitable organization activities become hidden actions that are unobserved or unverified by donors, creating a potential for moral hazard in the donation market.

The risk of potential charity fraud negatively impacts the donation behavior of principals. Previous economic research has reported that some donors care about the manner in which charitable funds are utilized.[1] For instance, Caviola et al. (2014) and Gneezy et al. (2014)

[1] This problem is recognized not only in economics research but also in general. A public opinion survey by the Japanese Cabinet Office revealed that 37% of individuals who choose not to donate to nonprofit organizations cite "difficulty in seeing the effect of donations" as a reason for their donation decision. Since

have revealed the phenomenon of overhead aversion, where donors seek to avoid their donations being used for overhead expenses by charitable organizations. This behavior suggests that donors are motivated by a desire to make a positive social impact with their donations. Karlan and Wood (2017) demonstrates that scientific information about the social impact of donations encourages donation behavior among those who have previously made significant donations. Moreover, Null (2011) and Metzger and Günther (2019) highlight the demand for information on the social impact of donations.[2] These studies suggest that charity frauds, which reduce the social impact of donations, negatively affect donation behavior.

We aim to understand the nature of charity fraud in terms of donation behavior and welfare of recipients by a laboratory experiment. We have developed a simple and unique game that incorporates elements of hidden action environment into the standard dictator game, in order to examine the impact of such environments on donation behavior and recipient welfare. In the standard dictator game, the dictator sends a portion of their initial endowment to the receiver. However, we introduce a twist to this game: the introduction of an intermediary (referred to as the deliverer) between the dictator and the receiver. The intermediary represents charitable organizations in real charity markets.

In this game, the dictator and deliverer make sequential decisions. First, the dictator sends a portion of their initial endowment to the deliverer. Then, the deliverer divides their initial endowment and the funds received from the dictator between themselves and the receiver.[3] Since the dictator's payoff does not depend on the deliverer's actions, in this game, the dictator cannot observe and infer the deliverer's behavior. Thus, the deliverer's actions become hidden actions. The differences with related games will be discussed in detail in Section 2.

We have two concrete research questions about nature of charity fraud. First, what

multiple answers are allowed, the possibility of self-serving bias is small. See the Cabinet Office's "Public Opinion Survey on NPOs" (conducted in 2013, in Japanese). https://survey.gov-online.go.jp/h25/h25-npo/index.html

[2] Duncan (2004) has developed theoretical models concerning donors who care about social impact, providing theoretical insights.

[3] More precisely, the deliverer decides the percentage of the funds received from the dictator to send to the receiver. The deliverer can choose from 0% to 200% in increments of 1%. If the deliverer chooses an allocation ratio higher than 100%, they transfer all the funds received from the dictator, along with a portion of their initial endowment to the receiver.

are the implications of the hidden action environment on the dictator's decision to give and the welfare of recipients after the transaction? To answer it, this study applied two treatments to this game. First, the dictator can specify their own actions as well as those of the deliverer (*enforcement* treatment). This treatment enables the dictator to observe and verify the deliverer's actions, eliminating the possibility of moral hazard. Second, the game described earlier is played as usual (*hidden action* treatment). This treatment entails the potential for moral hazard by the deliverer. By comparing these two treatments, we elucidate the manner in which asymmetric information situations affect donation behavior and recipient payoffs.

Second research question examines the breakdown of pledges on the agent side, which is typical of charity fraud.[4] To do so, we introduced a third treatment (*promise* treatment). Following Charness and Dufwenberg (2006) and Vanberg (2008), which investigated non-binding pre-play promises in a classical principle-agent relationship, before playing the game, the deliverer decides, in advance, how to share the pie with the receiver, and presents the same to the dictator. Then, similar to the hidden action treatment, the normal game is played. Therefore, the deliverer once again chooses the manner in which the pie is to be allocated, which determines the payoffs for each player.

The experiment's results reveal that the hidden action environment hinders the dictator's donation behavior and significantly decreases the receiver's welfare. Compared to the enforcement treatment, the hidden action treatment reduces the average donation amount by 35% for the dictator and decreases the receiver's average payoff by 85%.

Additionally, in a hidden action environment, non-binding promises by the deliverer offset the decrease in average donation amount to some extent. Nevertheless, the promises do not sufficiently improve the receiver's average payoff. Compared to the enforcement treatment, the promise treatment only reduces the average donation amount by 4% but still decreases the average payoff of the receiver by 71%. Furthermore, 56% of deliverers break their promises (choosing a lower share of the pie than promised). If all deliverers were to keep their promises, the average payoff for the receiver would double compared to the promise treatment. However, the amount would still not reach the average payoff without

[4] Some studies examine the effect of pledges on donors (principals) on altruistic behavior (see Andreoni and Serra-Garcia, 2021; Meyer and Tripodi, 2021).

information asymmetry. Taken together, contrary to Charness and Dufwenberg (2006) and Vanberg (2008), non-binding promises cannot fully resolve the negative consequences of the hidden action environment.

Our research complements studies on sequential contributions. When some donors are not aware of the quality of nonprofit activities, the contributions of early donors can serve as signals of this quality (Vesterlund, 2003; Andreoni, 2006). To examine this possibility, several studies have conducted games where multiple donors invest in public goods sequentially (Potters et al., 2005). The fact that the quality of donations is uncertain is a common starting point with our research. However, our focus differs from these studies. Sequential public goods games focus on the function of initial donations as signals in such environments, disregarding how the quality of donations is determined. Our research focuses on how the quality of donations depends on the manner in which charitable organizations utilize funds, and is determined under hidden action environment.[5]

Our research also contributes to the study of charity fraud. While empirical research on charity fraud in nonprofit organizations is quite limited, some studies have focused on social-economic attributes (Goel, 2020) and organizational characteristics (Gibelman and Gelman, 2001; McDonnell and Rutherford, 2018). These studies examine the factors that contribute to charity fraud under the assumption of information asymmetry. We quantitatively examine the negative impact of information asymmetry, thereby complementing a series of studies by once again demonstrating the importance of moral hazard in the charity market.[6]

The rest of the paper is structured as follows: In Section 2, we introduce the game used in our experiment and discuss the differences from related games. Section 3 provides an overview of the experiment. We present the experimental results in Section 4 and discuss them in Section 5. Finally, the conclusion is given in Section 6.

[5] The quality of donations in our experiment is represented by the deliverer's allocation ratio.

[6] Economic research on fraud is famously rooted in the study of credence goods initiated by Darby and Karni (1973). In credence goods markets, consumers do not know the quality of the goods they desire, while suppliers can ascertain the quality through diagnosis. Moreover, consumers can only observe the utility obtained from the traded goods (and may not even observe that). Dulleck et al. (2011) verifies the inefficiencies (under-treatment, over-treatment, overcharging) in credence goods markets through laboratory experiments. The charity market may differ slightly from typical credence goods markets. The quality of donations can be defined by the social impact of how much the donations have improved society. Charitable organizations may not easily discern the social impact truly desired by donors. Instead, donors may be the ones who know it.

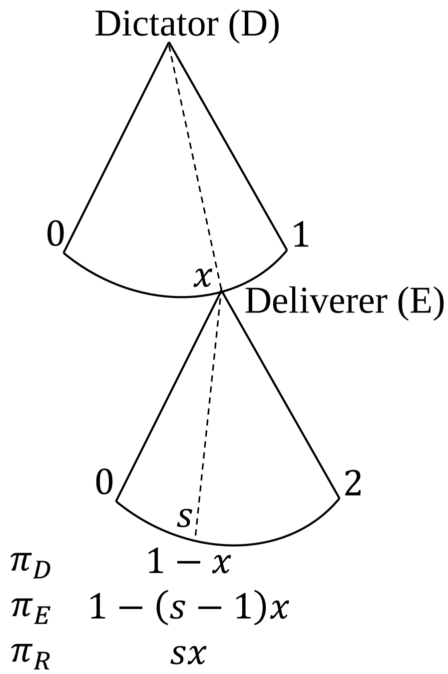

Figure 1: Game Tree. *Notes*: π_i is payoff of each player.

2 Dictator Game with Hidden Actions

2.1 The Game

Our game consists of a dictator (D), a deliverer (E), and a receiver (R). The dictator and deliverer have an initial endowment of I tokens, while the receiver has no initial endowment. The dictator and deliverer sequentially make decisions to determine the payoffs for all three players.

The dictator is the first mover, and decides how much of their initial endowment to send to the deliverer. Let the amount sent by the dictator be denoted as x. The second mover is the deliverer. The deliverer decides the percentage of tokens funded by the dictator to be sent to the receiver. The deliverer can choose a percentage ranging from 0% to 200% in 1% increments. Here, the allocation ratio chosen by the deliverer is denoted as $s \in \{0, 0.01, 0.02, \dots, 0.99, 1, 1.01, \dots, 1.98, 1.99, 2\}$.

The deliverer transfers sx tokens to the receiver, and this becomes the receiver's payoff.

If $s < 1$, the deliverer sends a portion of the sent amount from the dictator to the receiver, keeping the remaining $(1-s)x$ tokens as their own payoff. If $s > 1$, the deliverer adds the shortfall $(s-1)x$ tokens to the amount sent by the dictator from their initial endowment and sends it to the receiver. Therefore, the deliverer's payoff is $I - (s-1)x$ tokens. However, since the actions of the deliverer do not influence the dictator, the dictator's payoff is $I - x$ tokens. Our game tree is summarized in Figure 1.

2.2 Comparison with Related Games

2.2.1 Dictator Game

Our game represents a natural extension of the standard dictator game (e.g., Forsythe et al., 1994). In the dictator game, a player with an initial endowment chooses to transfer some of it to another player. In our game, when the deliverer transfers 100% of the dictator's transfer amount ($s = 1$) to the receiver, the final payoffs for the dictator and the receiver are $I - x$ and x tokens, respectively. These payoffs mirror the outcomes of the dictator game. Assuming a player who is only concerned with their own payoff, on the Nash equilibrium path of our game, the dictator does not send anything, and the deliverer chooses $s \in [0, 2]$. Therefore, in equilibrium, $\pi_D = I$, $\pi_E = I$, and $\pi_R = 0$, resembling the equilibrium of the dictator game. Thus, as in standard dictator games, the motive for dictators to send depends on pure/impure altruism (Andreoni, 1989, 1990; Andreoni and Miller, 2002) and distributional preferences (e.g., Fehr and Schmidt, 1999; Charness and Rabin, 2002).

Additionally, to examine rationality and preferences of altruistic behavior, some studies have conducted a dictator game with an exogenous price p in the form of a budget constraint, represented as $\pi_D + p\pi_R = I$, where π_D and π_R are the dictator's and receiver's payoffs, respectively (Andreoni and Miller, 2002; Fisman et al., 2007). In our game, the allocation ratio s of the deliverer can be interpreted as the price. The budget constraint for the dictator in our game is $\pi_D + (1/s)\pi_R = I$. In other words, as the allocation ratio increases, the cost of the transfer decreases.[7]

[7] An increase in the allocation ratio not only leads to a decrease in price but also introduces substitutability between the sent amount and the allocation ratio. Given s, the dictator maximizes $U(I - \pi_R/s, \pi_R)$ by choosing π_R. In this case, the optimal sent amount is $x^* = \pi_R^*/s$. Hence, as the allocation ratio increases, the sent amount decreases. Similarly, matching donations have both a positive effect due to price reduction

However, the standard dictator game does not reflect the hidden action environment in the charity market. Some studies have examined dictator games in which recipients are actual charitable organizations, thereby providing important insights into donation behavior.[8] Nevertheless, such games do not examine the extent to which recipient welfare improves through donations. By introducing an intermediary between the dictator and the recipient, our game can simultaneously assess donation behavior and recipient welfare, including moral hazard issues.

2.2.2 Trust Game

Our game is also related to trust games (Berg et al., 1995; Dufwenberg and Gneezy, 2000). In this game, the trustor entrusts a portion of their initial endowment to the trustee, which is then multiplied by a constant factor. The trustee decides the amount to be returned. In our game, the dictator (trustor) entrusts a portion of their initial endowment to the deliverer (trustee), who then shares $I + x$ tokens with the receiver. Thus, the dictator's transfer amount is multiplied by $(I + x)/x$ (a decreasing function of x) and entrusted to the deliverer. A significant difference from the trust game lies in who the trustee shares the benefits with. In the trust game, the trustee shares the benefits with the trustor. Therefore, a direct reciprocity towards the trustor is often observed in the trustee's behavior (Berg et al., 1995; Dufwenberg and Gneezy, 2000; McCabe et al., 2003). However, in our game, the deliverer shares the benefits with the receiver, and not with the dictator. So, direct reciprocity is not a motivating factor for the deliverer.[9]

Charness and Dufwenberg (2006) incorporates the essence of hidden action into the trust game. In this experiment, the trustor decides whether to send all of their initial

and a negative effect due to substitutability in individual contributions (Hungerman and Ottoni-Wilhelm, 2021). Thus, the deliverer's allocation ratio is a form of matching donation.

[8] Umer et al. (2022) conducted a recent meta-analysis on the dictator game, comprehensively examining various elements in the dictator game. See Umer et al. (2022) for the impact of recipient differences (students versus charitable organizations).

[9] Indirect reciprocity might be a motivating factor for the deliverer. This involves a person who has been treated kindly engaging in altruistic behavior towards an unrelated group. Strang et al. (2016) and Li et al. (2020) examine this motive, using a two-stage dictator game in which Player A shares a pie with Player B in the first stage, and then Player B shares another pie with Player C in the second stage. The results reveal that as Player A's transfer amount increases, Player B's transfer amount increases as well. In our game, as the dictator's transfer amount increases, the size of the pie shared between the deliverer and the receiver also increases. Therefore, deliverers with indirect reciprocity may increase the receiver's share of the pie.

endowment to the trustee. The trustee, in turn, decides to either (1) keep the entire amount sent by the trustor or (2) multiply the amount sent by a constant factor and return only the additional portion to the trustor. Importantly, when the trustee selects option (2), whether the amount sent is multiplied by a constant factor (resulting in additional return to the trustor) is probabilistic. As a result, the trustor cannot infer the trustee's actions from their own payoff. Thus, the trustee's choice becomes a hidden action, leading to potential moral hazard. Since the dictator's payoff does not depend on the deliverer's actions, in our game, even without introducing stochastic elements, the deliverer's choice constitutes a hidden action.

Charness and Dufwenberg (2006) and Vanberg (2008) demonstrated that in a hidden action environment, pre-communication (non-binding promises by the trustee) encourages trust behavior (the trustor sends money, and the trustee generates returns and sends them back to the trustor). In light of these findings, we investigate whether promises are effective in a hidden action environment in the charity market. Specifically, we examine whether non-binding promises by the deliverer promote the dictator's transfers and enhance receiver's payoffs.

2.3 Treatments

In order to examine the consequences of moral hazard in the charity market and the effects of non-binding promises, we used three treatment groups. The first two treatments focus on whether the deliverer's actions are hidden.

- *Enforcement* treatment: Stage 1: the dictator determines both x and s. Stage 2: the deliverer must choose s prescribed by the dictator.
- *Hidden action* treatment: Stage 1: the dictator decides x. Stage 2: the deliverer decides s.

In the enforcement treatment group, the dictator can observe and verify the allocation ratio. However, in the hidden action treatment group, the dictator cannot observe the allocation ratio. Thus, we compare the dictator's sent amount and receiver's payoff between the two treatment groups, in order to investigate the impact of moral hazard in the charity market on donation behavior and the welfare (payoff) of the receiver. Hoppe and Schmitz (2018) has also examined the effects of hidden action with similar treatments. The third treatment introduces promises from the deliverer in situations where moral hazard occurs.

- *Promise* treatment: Stage 0: the deliverer chooses a non-binding allocation ratio s_0 and discloses it to the dictator. Stage 1: the dictator determines x. Stage 2: the deliverer decides s. The payoffs of all three players are determined by s rather than s_0.

We compare the dictator's sent amount and receiver's payoff between the hidden action treatment and promise treatment groups. This comparison aims to examine the impact of the deliverer's promise on donation behavior and the welfare of the recipient in the presence of moral hazard in the charity market.

3 Experimental Design

In December 2023, we conducted experiments at the Research Institute for Socionetwork Strategies (RISS) at Kansai University, with participants from RISS's subject pool via ORSEE (Greiner, 2015). We obtained approval from the Institutional Review Board of the Research Institute for Socionetwork Strategies at Kansai University (Approval Number: 2023003).

The experiment consisted of eight sessions lasting 90 minutes each. Treatment assignments were as follows: initially, we evenly assigned the hidden action and promise treatment groups at the session level. Then, for sessions assigned to the hidden action treatment group, we assigned the enforcement treatment group to all groups in either odd or even rounds (assigning hidden action treatment group to the remaining rounds).[10]

The session procedure was as follows. First, participants took a quiz on the game structure and received explanations via a computer, in order to enhance their understanding. Next, participants played a one-shot game consisting of 14 rounds.[11] Groups were randomly

[10]Whether to assign enforcement treatment group in odd rounds was equally allocated at the session level.
[11]A calculator was provided to players during decision-making to predict the payoffs of all three players.

formed again, and roles were reassigned in each round. As an initial endowment, we provided 100 tokens to both the dictator and deliverer ($I = 100$). Additionally, participants were tasked with guessing each others' actions and beliefs.[12] The dictator decided the sent amount and guessed the deliverer's action (first-order belief). The deliverer determined the allocation ratio and guessed on the dictator's expectations regarding the deliverer's action (second-order belief). Meanwhile, the receiver guessed the deliverer's action when the dictator and deliverer were engaged in decision-making. After playing the game, participants took a survey on personal attributes and altruistic preferences. They did not receive feedback on game results after each round but could only review their own earnings in a summary table after survey. The experiment was computerized using oTree (Chen et al., 2016).[13]

Participants' experiment rewards consisted of participation fees (1,000 JPY) and performance fees (0 to 3,200 JPY) combined. Performance fees were determined as follows: for each participant, we randomly selected two rounds out of 14 and calculated the total tokens earned in both the game and the guess task, converting them at 5 JPY per token. Following Costa-Gomes et al. (2014), we adopted the quadratic scoring rule for the guess task for incentive compatibility.[14] Earnings in tokens were determined by $120 - 30((y - \hat{y})/100)^2$, where $\hat{y}$ is the predicted value, and y is the realized value of the target of prediction.[15]

The number of participants in the hidden action treatment and the enforcement treatment groups was 105 people (35 groups). Seven rounds were held per session for both the hidden action treatment and the enforcement treatment groups. Therefore, the total number of observations for the hidden action treatment and the enforcement treatment groups was 245. The number of participants in the promise treatment group was 90 people (30 groups).

[12] There was no guess task in the enforcement treatment group.

[13] You can find oTree source code and instructions at GitHub repository. See https://github.com/Kim-Kato-Lab/pdg-otree.

[14] Haaland et al. (2023) points out that with complex incentive structures, respondents may misunderstand, leading to the potential of not obtaining accurate reports. They suggest informing participants that "honest answers maximize earnings" and designing an interface where only interested participants can grasp the precise rules. We followed their suggestion and presented detailed scoring rule information on the web screen in a collapsible format with toggle buttons.

[15] Assume that participants believe the value of the predicted target Y follows a cumulative distribution $F(y)$. If the participant is risk-neutral, they would choose $\hat{y}$ to maximize the expected payoff of $120 - 30 \int ((y - \hat{y})/100)^2 dF(y)$. The first-order condition is $-(60/100)[\int y dF(x) - \hat{y}] = 0$. Therefore, the optimal response is $\hat{y} = \int y dF(x) = E(Y)$. In other words, a risk-neutral participant would answer with the expected value.

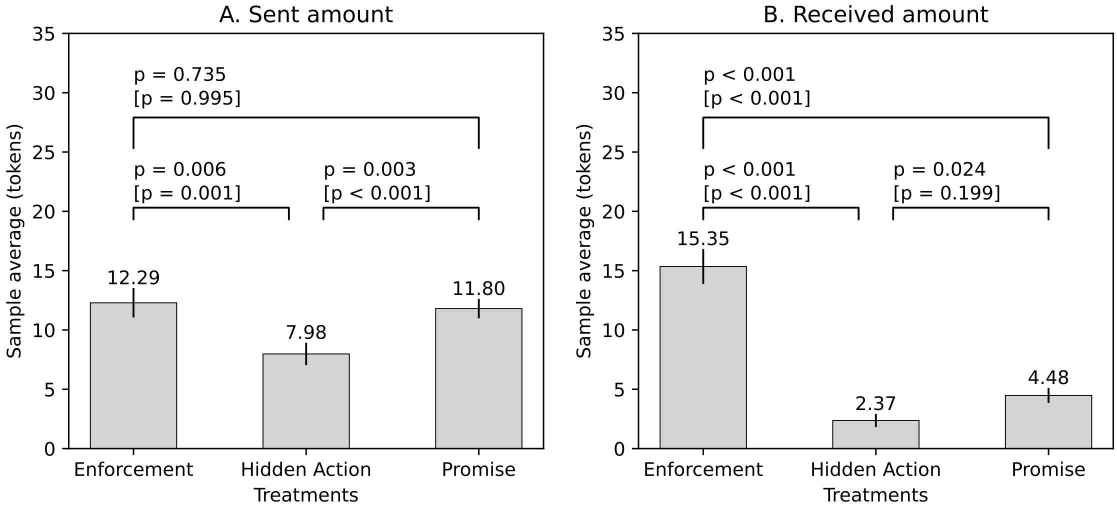

Figure 2: Average Sent Amount and Received Amount for Each Treatment Group. *Note*: Error bars represent standard errors of mean. The p-value is the result of a difference-in-means test between two treatment groups. The p-value in a square bracket is the result of Mann-Whitney U test.

There were 14 rounds per session for the promise treatment. Thus, the total number of observations for the promise treatment group was 420. The balance test between sessions is presented in Table 5 in Appendix A. As a result, because of a slight variation in students' academic fields across sessions, we controlled for round dummies and variables related to students' academic fields in the regression analysis.

4 Results

4.1 Sent Amount and Received Amount

Figure 2 showed the average amount sent by the dictator (left panel) and the average payoff of the receiver (right panel) in each treatment group. First, we examined the consequences of moral hazard in the charity market. Compared to the enforcement treatment group, the hidden action treatment decreases the average amount sent by 4.31 tokens, 35% reduction (7.98 tokens in the hidden action treatment group versus 12.29 tokens in the enforcement treatment group). Additionally, the hidden action treatment decreases the receiver's average payoffs by about 12.98 tokens, 85% reduction (2.37 tokens in the hidden action treatment group versus 15.35 tokens in the enforcement treatment group).[16] These differences are statistically significant and robust to controlling covariates and statistical inference with session-level clustered standard errors (see Table 6 in Appendix A). Furthermore, nonparametric tests such as the Mann-Whitney U test (Figure 2) and two-sample Kolmogorov-Smirnov test (Figure 5 in Appendix A) yield similar results.

Result 1: Hidden action environment discourages donation behavior and reduces recipients' welfare. The decrease in recipients' welfare exceeds the reduction in donations.

Next, we examined the effect of non-binding promises by the deliverer under asymmetric information. Compared to the hidden action treatment group, the promise treatment increases the average amount sent by about 3.82 tokens, 48% increase (7.98 tokens in the hidden action treatment group versus 11.8 tokens in the promise treatment group), which is statistically significant. Consequently, the average amount sent in the promise treatment group no longer exhibits a statistically significant difference compared with the enforcement treatment group. Compared with the enforcement treatment group, the promise treatment only reduces the average amount sent by 0.49 tokens (4% reduction). These results are robust to controlling covariates, using session-level clustered standard errors (Table 6 in Appendix A), and nonparametric tests (Figure 2 and Figure 5 in Appendix A). Thus, under asymmetric information conditions, introducing non-binding promises by the deliverer induces donation behavior comparable to situations without asymmetric information.

However, such promises do not sufficiently improve recipients' welfare. Compared with the hidden action treatment group, the promise treatment increases the receiver's average payoff by about 2.11 tokens, 89% increase (2.37 tokens in the hidden action treatment group versus 4.48 tokens in the promise treatment group), which is statistically significant by a simple t-test. However, this result is not robust to controlling covariates, using session-level clustered standard errors (Table 6 in Appendix A), and nonparametric tests (Figure 2 and Figure 5 in Appendix A). Compared with the enforcement treatment group, the promise treatment still reduces the receiver's average payoff by 10.87 tokens,

[16] Recent meta-analysis indicates that the average contribution of dictators (average payoff to receivers) in dictator games is 20% of the dictator's initial endowment (Umer et al., 2022). Since the enforcement treatment involves dictators deciding (x, s), there is no asymmetric information. In this case, receiver's average payoff is 15% of the dictator's initial endowment, which is slightly lower than in dictator games.

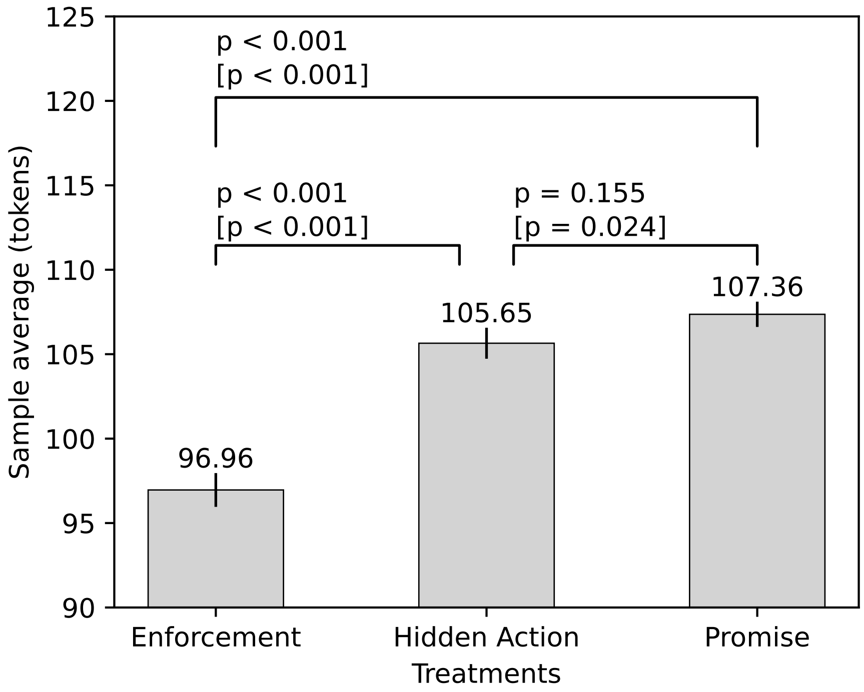

Figure 3: Average Deliverer's Payoff for Each Treatment Group. *Note*: Error bars represent standard errors of mean. The p-value is the result of a difference-in-means test between two treatments. The p-value in a square bracket is the result of Mann-Whitney U test.

comprising a decrease of 71%, which is statistically significant and robust. Therefore, the deliverer's non-binding promises do not sufficiently offset the negative effect on the welfare of the receiver in the hidden action environment.

Result 2: In a hidden action environment, non-binding promises by the deliverer can offset the negative effect of information asymmetry on the average amount sent. However, such promises do not sufficiently offset the negative effect on the receiver's average payoff.

Result 1 and 2 suggest that, regardless of the introduction of promises, hidden action environment provides opportunities for the deliverer to increase their payoffs. Figure 3 illustrates the average payoffs of the deliverer for each treatment group. Compared with to the enforcement treatment group, the hidden action treatment increases the average payoff by 9.69 tokens, 10% increase (105.65 tokens in the hidden action treatment group versus 96.96 tokens in the enforcement treatment group). Moreover, in comparison to the enforcement treatment group, the promise treatment increases the average payoff by 10.4 tokens, 11% increase (107.36 tokens in the promise treatment group versus 96.96 tokens in

Table 1: Response of Amount Sent

	Treatment: E		Treatment: HA and P	
	(1)	(2)	(3)	(4)
(Intercept)	10.155***	3.745	6.266***	2.597***
	(2.999)	(2.451)	(0.644)	(0.623)
Allocation ratio (percent)	0.019	-0.041*		
	(0.011)	(0.023)		
1 = Positive allocation ratio (percent)		16.975***		
		(4.872)		
First-order belief (percent)			0.110***	0.044
			(0.024)	(0.030)
1 = Positive first-order belief (percent)				10.459***
				(2.015)
Observations	245	245	665	665
Adjusted R^2	0.003	0.064	0.111	0.170

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. Outcome variable is the amount sent by the dictator (unit: token). In columns (1) and (2), the sample assigned to the enforcement treatment is used. In columns (3) and (4), the sample assigned the hidden action or promise treatment is used.

the enforcement treatment group). These differences are statistically significant and robust to controlling covariates, using clustered standard errors (see Table 7 in Appendix A), and nonparametric tests (see Figure 6 in Appendix A for two-sample KS test). Therefore, the deliverers can leverage asymmetric information to enhance their own payoffs.

4.2 Why Does the Hidden Action Treatment Reduce Sent Amount?

Hidden action environments decrease the average amount sent by dictators, but the introduction of promises offsets this negative effect. In order to examine this result in detail, we first investigate the relationship between the amount sent and allocation ratio, using samples from the enforcement treatment group where dictators decide both these values. Columns (1) and (2) of Table 1 present the estimated results. Column (1) suggests no correlation between the amount sent and the allocation ratio. However, column (2) implies that the lack of correlation is a result of the two conflicting effects. First, when dictators decide to send their own contribution to the receiver ($s > 0$), they significantly increase the amount sent. If dictators care about the deliverer's payoff, the coefficient of the dummy variable, indicating a positive allocation ratio, should be negative. This reveals that dictators prioritize the receiver's payoff over the that of the deliverer's. Second, while the statistical significance is marginal, when dictators decide to send a large portion of their own contribution to the receiver (large s), they may simultaneously decrease the amount sent.

Similar behavioral responses are observed even in hidden action environments. Columns (3) and (4) of Table 1 present the estimated results using samples from the hidden action treatment group and the promise treatment group. Notably, regardless of the presence of promises, dictators cannot observe the allocation ratio determining the deliverer's and receiver's payoffs in hidden action environments. Therefore, we use the dictator's expectations about the allocation ratio (first-order belief) as explanatory variables for the same . Column (3) indicates a positive correlation between the amount sent and the first-order belief. In particular, this correlation is due to whether dictators expect a positive allocation ratio. By column (4), dictators significantly increase the amount sent when dictators expect some of their own transfers to contribute to the receiver's payoff. The motives behind such behavioral responses are discussed in Section 5.1.

Next, based on the dictator's behavioral responses, we explain the reason behind the hidden action treatment reducing the average amount sent. Again, dictators significantly increase the amount sent when they decide (expect) that some of their contribution will go to the receiver. In the enforcement treatment group, 78% of dictators choose positive allocation ratio ($s > 0$). In the hidden action treatment group, 48% of dictators expect that some of their contribution will be allocated to the receiver's payoff ($E_D(s) > 0$).[17] Therefore, information asymmetry reduces the proportion of dictators who anticipate that some of their contributions will benefit the receiver at the decision-making stage, resulting in the negative effect of the hidden action treatment. The effect of the hidden action treatment via this mechanism is $-8.22 (= 10.459 \times 0.48 - 16.975 \times 0.78)$ tokens, which is greater than the overall effect (-4.31 tokens). This is because the impact of the substitutability between x and s in the enforcement treatment group is not considered.[18] Taking this into account, the effect of the hidden action treatment becomes -3.55 tokens, which is closer to the overall effect.

[17] The difference between the proportion of $s > 0$ in the enforcement treatment group and the proportion of $E_D(s) > 0$ in the hidden action treatment group is statistically significant (t-test and Mann-Whitney U test: $p < 0.001$).

[18] The average allocation ratio in the enforcement treatment is 114%. Due to the substitutability between x and s, the amount sent in the enforcement treatment decreased by 4.67 tokens.

Table 2: Decompose of Average Payoff of Player R

	$Cov(x, s)$	$E(x)E(s)$	Average payoff of receivers
Enforcement	1.38	13.98	15.35
Hidden action	-0.76	3.13	2.37
Promise	-0.95	5.43	4.48

In the promise treatment group, 66% of dictators expect some of their contribution to benefit the receiver's payoff.[19] Therefore, under the information asymmetry, the introduction of promises enhances such expectations, leading to a higher average contribution in the promise treatment group compared with the hidden action treatment group. The effect of the promise treatment via this expectation formation is $1.883 (= 10.459 \times (0.66 - 0.48))$ tokens, representing half of the overall effect (3.82 tokens).

4.3 Why Does Not Promise Improve Recipient's Welfare?

As stated in Result 2, despite offsetting the decrease in average amount sent because of information asymmetry, non-binding promises do not sufficiently improve the receiver's payoff. To examine the underlying cause of this result, we decompose the receiver's average payoff using the covariance formula into two parts: $E(\pi_R) = \text{Cov}(x, s) + E(x)E(s)$. The results are presented in Table 2. The first component is the covariance between the amount sent and allocation ratio. This reflects the increase in the marginal effect of the amount sent on the recipient's payoff, as the allocation ratio increases. While the covariance in the enforcement treatment group is positive, the covariance is negative in the hidden action and promise treatment groups. Through this component, the promise treatment reduces the recipient's average payoff by about $2.33 (= -0.95 - 1.38)$ tokens compared to the enforcement treatment (21% of the overall negative effect of the promise treatment).

The positive covariance in the enforcement treatment group arises from the significant increase in contributions by dictators who choose a positive allocation ratio (as observed in column (2) of Table 1). In fact, restricting samples where dictators who choose a positive allocation ratio yields a negative covariance (-1.9) in the enforcement treatment group. However, as shown later, behavioral responses in the hidden action and promise

[19] The proportion of $E_D(s) > 0$ differs significantly between the hidden action treatment and the promise treatment groups (t-test and Mann-Whitney U test: $p < 0.001$).

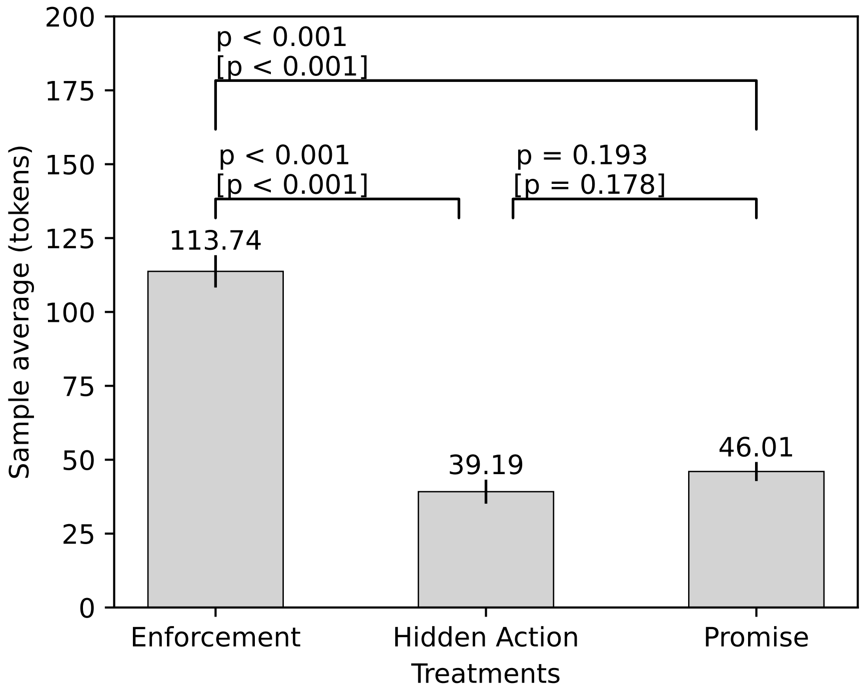

Figure 4: Average Allocation Ratio for Each Treatment Group. *Note*: Error bars represent standard errors of mean. The p-value is the result of a difference-in-means test between two treatments. The p-value in a square bracket is the result of Mann-Whitney U test.

treatment groups do not involve a significant increase in the allocation ratio due to positive amounts sent. Instead, a strong behavioral response is observed, in which the allocation ratio decreases as the amounts sent increases. Consequently, the covariances in the hidden action and promise treatment groups are negative.[20]

The second element is the product of the average amount sent and the average allocation ratio. This simply reflects that increases in both the average amount sent and average allocation ratio increase the receiver's average payoff. The average allocation ratios for each treatment group are shown in Figure 4. As a result, compared with the enforcement treatment group, the average allocation ratio in the hidden action treatment group is 74.55 percentage points lower, and in the promise treatment group, it is 67.73 percentage points lower. However, there is no statistically significant difference in the average allocation ratio between the promise and hidden action treatment groups. These differences are statistically significant and robust to controlling covariates and dictator's contributions,

[20]Figure 8 in Appendix A shows joint distribution of amount sent and allocation ratio for each treatment.

Table 3: Response of Allocation Ratio

	Hidden action treatment	Promise treatment	
	(1)	(2)	(3)
(Intercept)	26.226***	35.090***	27.884***
	(4.383)	(5.481)	(5.538)
Amount sent (token)	-0.483***	-0.463**	-0.696***
	(0.136)	(0.230)	(0.257)
Second-order belief (percent)	0.427***		0.616***
	(0.028)		(0.104)
Promise of allocation ratio (percent)		0.238***	0.024
		(0.071)	(0.033)
Observations	245	420	420
Adjusted R^2	0.129	0.049	0.232

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. Outcome variable is allocation ratio (unit: percent). In column (1), we use the sample from the hidden action treatment. In columns (2) and (3), we use the sample from the promise treatment.

cluster standard errors, and nonparametric tests.[21] Therefore, non-binding promises fail to sufficiently increase the allocation ratio, leading to no improvement in the receiver's average payoff. As a result, the promise treatment reduces the receiver's average payoff by $-8.5(= 4.48 - 15.35)$ tokens, accounting for 79% of the overall negative effect of the promise treatment group.

The reason for the higher allocation ratio in the enforcement treatment group lies in the differences in decision-makers regarding the allocation ratio. In the enforcement treatment group, dictators determine the allocation ratio. Since the allocation ratio does not affect dictators' payoffs at all, altruistic dictators have an incentive to maximize the allocation ratio as much as possible. On the other hand, in the hidden action and promise treatment groups, the deliverer determines the allocation ratio. Altruistic deliverers face a trade-off between the decrease in their payoff (π_E) and the increase in the receiver's payoff (π_R). As a result, the average allocation ratio in the enforcement treatment group is higher than that in the hidden action and promise treatment groups.

Next, we estimate the deliverer's behavioral response to examine why no difference is observed in the average allocation ratio between the hidden action and promise treatment groups. The estimation results of the response of the allocation ratio to the amount sent and the second-order belief (deliverer's expectation regarding the dictator's expectation

[21] Regression results are presented in Table 8 in Appendix A. Additionally, Figure 7 in Appendix A shows the distribution of allocation ratios and the results of the two-sample Kolmogorov-Smirnov test.

about the allocation ratio) are presented in Table 3. Columns (1) and (3) indicate that as dictators increase their contributions, deliverers decrease the allocation ratio. Through this mechanism, the promise treatment decreases the allocation ratio by $4.36(= -0.696 \times 11.80 - (-0.483) \times 7.98)$ percentage points.

The higher the second-order belief, the more deliverers increase the allocation ratio. The promise treatment does not increase the average second-order belief (40.06% in the promise treatment group versus 39.36% in the hidden action treatment group). However, the coefficient value for the second-order belief may differ between the two treatments (z-score: 1.755, p-value: 0.079). This suggests that the promise treatment may exhibit a more sensitive response to the second-order belief. Through this mechanism, the promise treatment increases the allocation ratio by $7.87(= 0.616 \times 0.4006 - 0.427 \times 0.3936)$ percentage points.

Column (2) shows that the increase in the allocation ratio at the promise stage increases the allocation ratio actually chosen. From columns (3) we note that once the second-order belief is controlled for, this correlation disappears, suggesting that the effect of the allocation ratio at the promise stage only influences the allocation ratio chosen through the second-order belief.

In summary, the promise treatment has both a negative effect from the substitution with the amount sent and a positive effect from the sensitive response to the second-order belief. The motives behind such behavioral responses are discussed in Section 5.2. The cancellation of these conflicting effects results in no difference in the average allocation ratio between the two treatment groups.[22]

4.4 Welfare Loss for Recepients by Breaking Promises

As promises are non-binding, the deliverer can choose a lower allocation ratio than that proposed during the promise stage and break the promise. 56% of deliverer actually break their promises. Table 4 estimates the propensity of deliverer to break promises. The higher the proposed distribution ratio in the promise stage, the more likely the deliverer is to break

[22] The sum of these conflicting effects is 3.514% percentage points, accounting for half of the overall effect.

Table 4: Regressions of Response of Deliverer

	1 = Breaking promises
	(1)
(Intercept)	0.497***
	(0.025)
Allocation ratio in the promise stage (percent)	0.004***
	(0.001)
Second-order belief (percent)	-0.004***
	(0.001)
Amount sent (token)	0.003**
	(0.001)
Observations	256
Adjusted R^2	0.196

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. We use sample with $x > 0$ in the promise treatment. For interpretation purposes, the explanatory variable "Amount sent" is adjusted so that the smallest positive contribution (one token) is zero.

their promise. Those who keep their promises propose an average distribution ratio of 48.3% in the promise stage and actually choose an average distribution ratio of 88.5%. In contrast, deliverers who break their promises propose an average distribution ratio of 85.1% in the promise stage and actually choose an average distribution ratio of 12.6%. Moreover, the deliverer is more likely to break their promises when the dictator allocates a larger amount, or when second-order beliefs are higher. This corresponds to the deliverers' behavioral responses seen in Table 3.

If deliverers were to actually choose the allocation ratio proposed in the promise stage, receiver's average payoff would increase from 4.48 tokens to 9.59 tokens (a 114% increase). Therefore, breaking promises by deliverer results in a reduction of receiver's payoff to less than half. By the t-tests, the counterfactual average payoff of receiver if all promises were kept is statistically significantly higher than the average payoff in the hidden action treatment group ($p < 0.001$), and statistically significantly lower than the average payoff in the enforcement treatment group ($p = 0.001$). This suggests that even if everyone were to keep their promises, the introduction of promises cannot cover the welfare loss of receiver due to information asymmetry.

5 Discussion

5.1 Dictator's Motivation

As we demonstrate in Section 4.2, dictators significantly increase the amount sent when they anticipate (or decide) that a portion of their contribution will improve the receiver's payoff. This finding is consistent with that of Ashraf et al. (2006), where trustors are more likely to exhibit trust behavior when the expected return in trust games is high. The dictator's contributions stem from altruistic motives to enhance the receiver's payoff. This motive can be explained by various models, including the pure altruism model (Andreoni and Miller, 2002), impact philanthropy model considering the social impact of donations (Duncan, 2004), and social preference models incorporating the maximin criterion, considering the player with the lowest payoff (Charness and Rabin, 2002).[23] Therefore, dictators are more likely to exhibit trust behavior towards the deliverer, significantly increasing the amount sent, when they anticipate that the deliverer will engage in altruistic behavior, i.e., when dictators expect high altruistic utility.

5.2 Deliverer's Motivation

5.2.1 Guilt Aversion

As shown in Section 4.3, the deliverer increases the allocation ratio proportionally to the second-order belief. This is akin to the findings of Dufwenberg and Gneezy (2000), wherein the trustee is more likely to exhibit trustworthy behavior, as they believe that the trustor expects them to act trustworthily in trust games. Dufwenberg and Gneezy (2000) interprets this result as the trustee's motivation of guilt aversion, where they do not want to return a payoff that falls below the trustor's expectations, as formalized by Charness and Dufwenberg (2006). However, interpreting the deliverer's behavioral response in terms of the original guilt aversion is challenging since the payoff of the dictator, who is the trustor,

[23] The inequality aversion model (Fehr and Schmidt, 1999) is another representative example of outcome-based social preference models. According to this model, dictators are motivated to reduce inequality between themselves and the receiver (inequality advantageous to the dictator) without widening inequality between the dictator and the deliverer (inequality disadvantageous to the dictator). Details on the theoretical model are provided in the Appendix B.

does not depend on the deliverer's actions.

As mentioned earlier, the dictator sends funds to the deliverer not because they expect an increase in their own payoff, but because they anticipate an improvement in the receiver's payoff. Therefore, if the deliverer perceives the dictator's motivation in this way, they would feel guilty about taking actions that fall below the dictator's expectations regarding the receiver's payoff. In our game, the deliverer's guilt aversion entails avoiding choosing allocation ratios that fall below the dictator's expected altruistic utilities. This type of guilt aversion suggests a positive correlation between the allocation ratio and second-order belief. Moreover, the positive correlation in the promise treatment group may be stronger than in the hidden action treatment group. This could imply that even in the absence of binding forces, promises may evoke guilt.[24]

5.2.2 Altruistic Preferences

A substitution relationship is observed between the dictator's transfer amount and deliverer's allocation ratio. This is suggestive of the deliverer's altruistic motive to increase the receiver's payoff, while also avoiding guilt. Here, consider a CES-type pure altruism model. Andreoni and Miller (2002) shows that this utility function best explains the outcomes of the dictator game. This model can be expressed as $U_E = ((1-\beta)\pi_E^{\rho} + \beta\pi_R^{\rho})^{1/\rho}$. The parameter β represents altruism, and the parameter ρ indicates substitutability. A deliverer with this utility function determines the optimal share ratio of the pie $p = \pi_R/(1+x)$ based only on β and ρ.[25] An increase in the transfer amount by the dictator increases the share ratio p of the pie.[26] Therefore, the deliverer reduces the allocation ratio to maintain the optimal share ratio of the pie.

[24]It is important to note the sources of guilt aversion. There are two types of guilt aversion induced by promises. First, individuals feel guilt due to deviations from the trustee's expectations of the trustor's beliefs (second-order belief). Charness and Dufwenberg (2006) demonstrates that promises induce guilt by increasing second-order belief. Second, individuals feel guilty due to deviations from the obligation of "what one should do." Vanberg (2008) shows that promises themselves entail such obligations and evoke guilt. The positive correlation between the allocation ratio and second-order belief is consistent with the first source. However, unlike Charness and Dufwenberg (2006), our experimental results do not show that the promise treatment increases second-order belief. The result that the promise treatment is more sensitive to second-order belief might relate to the second source.

[25]When the dictator makes a constant contribution, the deliverer divides the pie $1 + x$, which is the sum of their initial endowment and the dictator's contribution, with the receiver. The share ratio p of the pie indicates the receiver's percentage. See C for detailed derivation.

[26]$\partial p/\partial x = s/(1+x)^2 > 0$

The substitutability between the dictator's transfers and the deliverer's allocation ratio is also justified in other outcome-based social preference models. Fehr and Schmidt (1999)'s inequality aversion model suggests that the optimal allocation ratio is one where $\pi_R = \pi_E$. Similarly, Charness and Rabin (2002)'s social preference model proposes that the optimal allocation ratio is one where $\pi_R = \pi_D$. These optimal allocation ratios are decreasing functions of x.[27] However, few deliverers behave consistently with these models. Table 9 in Appendix A categorizes the motives of deliverers based on outcomes. In both the hidden action and promise treatment groups, in around 8% of cases, deliverers achieve equal outcomes such as $\pi_R = \pi_E$ or $\pi_R = \pi_D$ by choosing $s \in (0, 2)$. In such cases, around 23-25% of altruistic deliverers exhibit inequality aversion or care for the maximin criterion ($2/8$ in the hidden action treatment group; $5/16$ in the promise treatment group).[28]

5.3 Implications to Charity Market

Deliverers face a trade-off, according to which increasing the allocation ratio to enhance the receiver's payoff may diminish their own potential payoff. In other words, a conflict of interest exists between the dictator and the deliverer. Despite motivations to avoid guilty and to enhance the receiver's payoff, the hidden action environment provides deliverers with a strong incentive to act selfishly. Dictators anticipate this and hesitate to contribute in such environments. In a market structure where donors cannot observe and verify the social impact of their donations, donors may perceive their contributions as having a weak impact on society. As a result, the hidden action environment significantly decreases the welfare of the receiver more than the attenuation of donation behavior. Considering the inherent purpose of donations to improve society, this result suggests that the moral hazard issue in the donation market is more severe than anticipated.

To encourage donation behavior in hidden action environment, interventions that upwardly revise beliefs about social impact are effective. For instance, information provision

[27] The allocation ratio that satisfies $\pi_R = \pi_E$ is $s^* = (1 + x)/2x$, and $ds^*/dx = -1/2x^2 < 0$. Similarly, the allocation ratio that satisfies $\pi_R = \pi_D$ is $s^* = (1 - x)/x$, and $ds^*/dx = -1/x^2 < 0$. See Appendix C for detailed derivation.

[28] Although we provide calculator in the experiment, it is quite challenging for deliverers to find the allocation ratio that achieves equal outcomes. Considering this cognitive limitation, the proportion of deliverers inferred from observational data to exhibit inequality aversion or care for the maximin criterion should be interpreted as a lower bound.

regarding social impact, as examined in Karlan and Wood (2017), is one such intervention. Additionally, institutional mechanisms where charitable organizations outline their activities in advance (such as non-binding promises by deliverers in our experiment) constitute another approach. In our experiment, such promises significantly improved donors' expectations regarding social impact and effectively encouraged donation behavior, even in a hidden action environment. However, non-binding promises did not increase deliverers' allocation ratios, or sufficiently improve the receiver's welfare.

The greatest improvement in the welfare of the receiver takes place when the dictator is able to determine the allocation ratio. This is similar to specific donation systems, where donors can specify the use of their donations while determining the amount. Such systems are commonly utilized in universities and research institutes. However, this involves several considerations. First, there remains a risk of charitable organizations misusing specific funds. Nevertheless, compared to general donations, specific donation systems undoubtedly reduce the hidden actions of charitable organizations. Moreover, this system may make it easier for donors to observe and verify the social impact of their donations. Our experiment ensures that deliverers adhere to the allocation ratios specified by donors, leaving no room for hidden actions.

Second, the productivity of uses specified by donors may be lower than those determined by charitable organizations. Generally, charitable organizations are expected to understand the support needed better than donors. In such cases, donors should delegate the use of donations to charitable organizations. Our experiment provides common knowledge about the receiver's environment and impact of deliverer's behavior on the receiver's payoff. Therefore, in terms of information advantage, delegating the handling of contributions to charitable organizations is not necessarily appropriate in our experiment.

Two limitations regarding our experiment have been identified. The first pertains to the structure of the receiver's payoff. Altruistic deliverers determine the allocation ratio as inversely proportional to dictator's contributions. This is because the receiver's payoff is determined by the product of the dictator's contribution and deliverer's allocation ratio. Therefore, this payoff structure might significantly decrease the receiver's payoff. As such, reexamining the structure of the receiver's payoff to ensure that the altruistic behavior of deliverers does not substitute for that of the dictator is necessary. The conclusion regarding the decrease in the receiver's payoff might be conservative.

The second limitation is with regard to the deliverer's motives. Since we incentivize their earnings in the game, deliverers face the trade-off mentioned earlier. While criminals attempting charity fraud or for-profit organizations operating charities (as discussed by Rose-Ackerman (1996)) may have selfish motives, it is unlikely that most nonprofit organizations dominating the donation market harbor such strong selfish motives. Nonprofit organizations produce public goods that are not provided by governments or markets, and support those left behind by government aid or market mechanisms. Therefore, to bring the experiment closer to real markets, economic incentives of the deliverers must be tied to the payoffs of the receivers. Nonetheless, the hidden action environment suggests opportunities for profit-seeking organizations or criminals to act more selfishly. In particular, the result that half of the deliverers break promises and engage in selfish behavior suggests that the hidden action environment breeds charity fraud.

6 Conclusions

This study created a game incorporating the principal-agent relationship into the dictator game, in order to examine the consequences of moral hazard in the donation market. The hidden action environment in the donation market not only hampers donors' donation behavior, but also potentially further harms the recipient's welfare. Introducing non-binding promises in such an environment promotes donation behavior. However, exploiting this, fundraising organizations have the opportunity to enhance their own payoffs, with little improvement in recipient welfare. As future tasks, it is necessary to examine the effectiveness of monitoring systems and institutions in which charities specify and adhere to their operational policies in advance (binding commitments).

References

Andreoni, J. (1989), 'Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence', *Journal of Political Economy* **97**(6), 1447–1458.

Andreoni, J. (1990), 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving', *The Economic Journal* **100**(401), 464.

Andreoni, J. (2006), 'Leadership Giving in Charitable Fund-Raising', *Journal of Public Economic Theory* **8**(1), 1–22.

Andreoni, J. and Miller, J. (2002), 'Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism', *Econometrica* **70**(2), 737–753.

Andreoni, J. and Serra-Garcia, M. (2021), 'The Pledging Puzzle: How Can Revocable Promises Increase Charitable Giving?', *Management Science* **67**(10), 6198–6210.

Archambeault, D. S., Webber, S. and Greenlee, J. (2015), 'Fraud and Corruption in U.S. Nonprofit Entities: A Summary of Press Reports 2008-2011', *Nonprofit and Voluntary Sector Quarterly* **44**(6), 1194–1224.

Ashraf, N., Bohnet, I. and Piankov, N. (2006), 'Decomposing trust and trustworthiness', *Experimental Economics* **9**(3), 193–208.

Berg, J., Dickhaut, J. and McCabe, K. (1995), 'Trust, Reciprocity, and Social History', *Games and Economic Behavior* **10**(1), 122–142.

Caviola, L., Faulmüller, N., Everett, Jim. A. C., Savulescu, J. and Kahane, G. (2014), 'The evaluability bias in charitable giving: Saving administration costs or saving lives?', *Judgment and Decision Making* **9**(4), 303–315.

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

Charness, G. and Rabin, M. (2002), 'Understanding Social Preferences with Simple Tests', *The Quarterly Journal of Economics* **117**(3), 817–869.

Chen, D. L., Schonger, M. and Wickens, C. (2016), ‘oTree—An open-source platform for laboratory, online, and field experiments’, *Journal of Behavioral and Experimental Finance* **9**, 88–97.

Costa-Gomes, M. A., Huck, S. and Weizsäcker, G. (2014), ‘Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect’, *Games and Economic Behavior* **88**, 298–309.

Darby, M. R. and Karni, E. (1973), ‘Free Competition and the Optimal Amount of Fraud’, *The Journal of Law and Economics* **16**(1), 67–88.

Dufwenberg, M. and Gneezy, U. (2000), ‘Measuring Beliefs in an Experimental Lost Wallet Game’, *Games and Economic Behavior* **30**(2), 163–182.

Dulleck, U., Kerschbamer, R. and Sutter, M. (2011), ‘The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition’, *American Economic Review* **101**(2), 526–555.

Duncan, B. (2004), ‘A theory of impact philanthropy’, *Journal of Public Economics* **88**(9-10), 2159–2180.

Fehr, E. and Schmidt, K. M. (1999), ‘A Theory of Fairness, Competition, and Cooperation’, *The Quarterly Journal of Economics* **114**(3), 817–868.

Fisman, R., Kariv, S. and Markovits, D. (2007), ‘Individual Preferences for Giving’, *American Economic Review* **97**(5), 1858–1876.

Forsythe, R., Horowitz, J. L., Savin, N. and Sefton, M. (1994), ‘Fairness in Simple Bargaining Experiments’, *Games and Economic Behavior* **6**(3), 347–369.

Gibelman, M. and Gelman, S. R. (2001), ‘Very Public Scandals: Nongovernmental Organizations in Trouble’, *Voluntas: International Journal of Voluntary and Nonprofit Organizations* **12**(1), 49–66.

Gneezy, U., Keenan, E. A. and Gneezy, A. (2014), ‘Avoiding overhead aversion in charity’, *Science* **346**(6209), 632–635.

Goel, R. K. (2020), ‘Uncharitable Acts in Charity: Socioeconomic Drivers of Charity-Related Fraud’, *Social Science Quarterly* **101**(4), 1397–1412.

Greiner, B. (2015), ‘Subject pool recruitment procedures: Organizing experiments with ORSEE’, *Journal of the Economic Science Association* **1**(1), 114–125.

Haaland, I., Roth, C. and Wohlfart, J. (2023), ‘Designing Information Provision Experiments’, *Journal of Economic Literature* **61**(1), 3–40.

Hoppe, E. I. and Schmitz, P. W. (2018), ‘Hidden action and outcome contractibility: An experimental test of moral hazard theory’, *Games and Economic Behavior* **109**, 544–564.

Hungerman, D. M. and Ottoni-Wilhelm, M. (2021), ‘Impure Impact Giving: Theory and Evidence’, *Journal of Political Economy* **129**(5), 1553–1614.

Jensen, M. C. and Meckling, W. H. (1976), ‘Theory of the firm: Managerial behavior, agency costs and ownership structure’, *Journal of Financial Economics* **3**(4), 305–360.

Karlan, D. and Wood, D. H. (2017), ‘The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment’, *Journal of Behavioral and Experimental Economics* **66**, 1–8.

Li, J., Cheo, R. and Xiao, E. (2020), ‘The effect of voice on indirect reciprocity: Results from the lab’, *Economics Letters* **189**, 108988.

McCabe, K. A., Rigdon, M. L. and Smith, V. L. (2003), ‘Positive reciprocity and intentions in trust games’, *Journal of Economic Behavior & Organization* **52**(2), 267–275.

McDonnell, D. and Rutherford, A. C. (2018), ‘The Determinants of Charity Misconduct’, *Nonprofit and Voluntary Sector Quarterly* **47**(1), 107–125.

Metzger, L. and Günther, I. (2019), ‘Making an impact? The relevance of information on aid effectiveness for charitable giving. A laboratory experiment’, *Journal of Development Economics* **136**, 18–33.

Meyer, C. J. and Tripodi, E. (2021), ‘Image concerns in pledges to give blood: Evidence from a field experiment’, *Journal of Economic Psychology* **87**, 102434.

Null, C. (2011), 'Warm glow, information, and inefficient charitable giving', *Journal of Public Economics* **95**(5-6), 455–465.

Potters, J., Sefton, M. and Vesterlund, L. (2005), 'After you—endogenous sequencing in voluntary contribution games', *Journal of Public Economics* **89**(8), 1399–1419.

Rose-Ackerman, S. (1996), 'Altruism, Nonprofits, and Economic Theory', *Journal of Economic Literature* **34**(2), 701–728.

Strang, S., Grote, X., Kuss, K., Park, S. Q. and Weber, B. (2016), 'Generalized Negative Reciprocity in the Dictator Game – How to Interrupt the Chain of Unfairness', *Scientific Reports* **6**(1), 22316.

Umer, H., Kurosaki, T. and Iwasaki, I. (2022), 'Unearned Endowment and Charity Recipient Lead to Higher Donations: A Meta-Analysis of the Dictator Game Lab Experiments', *Journal of Behavioral and Experimental Economics* **97**, 101827.

Vanberg, C. (2008), 'Why Do People Keep Their Promises? An Experimental Test of Two Explanations', *Econometrica* **76**(6), 1467–1480.

Vesterlund, L. (2003), 'The informational value of sequential fundraising', *Journal of Public Economics* **87**(3-4), 627–657.

Appendix

A Figures and Tables

Table 5: Balance Test

	HA & Enforcement	Promise	F-test, p-value
1 = Male	0.39	0.41	0.85
1 = Foreigner	0.04	0.04	0.90
Age	20.83	20.70	0.56
Academic year	2.86	2.74	0.40
1 = Social science	0.59	0.69	0.02
1 = Economics	0.21	0.16	0.45
1 = Humanities	0.22	0.23	0.86
1 = Natural science	0.19	0.08	0.11
1 = Experience with lab experiments	0.88	0.90	0.71

Note: Average of session-level average of participant's characteristics. The HA & Enforcement column contains the sessions that are assigned the hidden action treatment and the enforcement treatment. The Promise column contains sessions that are assigned the promise treatment. F-test examines the null hypothesis HA & Enforcement = Promise.

Table 6: Regressions of Sent Amount and Received Amount

	Sent amount		Received amount	
	(1)	(2)	(3)	(4)
Hidden action	-4.310***	-4.249***	-12.978***	-12.908***
	(1.444)	(0.869)	(2.783)	(1.806)
Promise	-0.485	-0.383	-10.870***	-10.625***
	(2.255)	(2.021)	(3.214)	(2.725)
Control average	12.290	12.290	15.349	15.349
F-test, p-value				
Hidden action = Promise	0.030	0.033	0.207	0.151
Covariates		X		X
Observations	910	910	910	910
Adjusted R^2	0.009	0.036	0.100	0.115

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. Covariates are round fixed effects, patron's and dictator's academic fields (social science dummy, economics dummy, and natural science dummy).

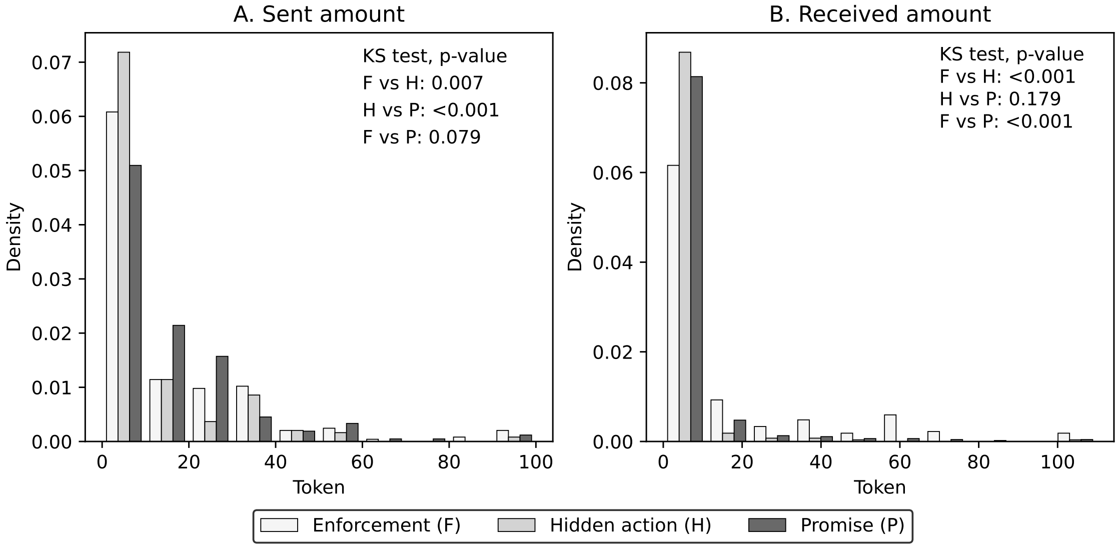

Figure 5: Distribution of Amount Sent by Treatment Group. *Note*: We show the p-value of the two-sample Kolmogorov–Smirnov test in the upper right corner of each panel.

Table 7: Regressions of Dictator's Payoff

	Payoff of Deliverer (unit: token)	
	(1)	(2)
Hidden action	8.690***	8.680***
	(1.643)	(1.340)
Promise	10.403***	10.255***
	(1.495)	(1.189)
Control average	96.959	96.959
F-test, p-value		
Hidden action = Promise	0.048	0.025
Covariates		X
Observations	910	910
Adjusted R^2	0.076	0.079

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. Covariates are round fixed effects, patron's and dictator's academic fields (social science dummy, economics dummy, and natural science dummy).

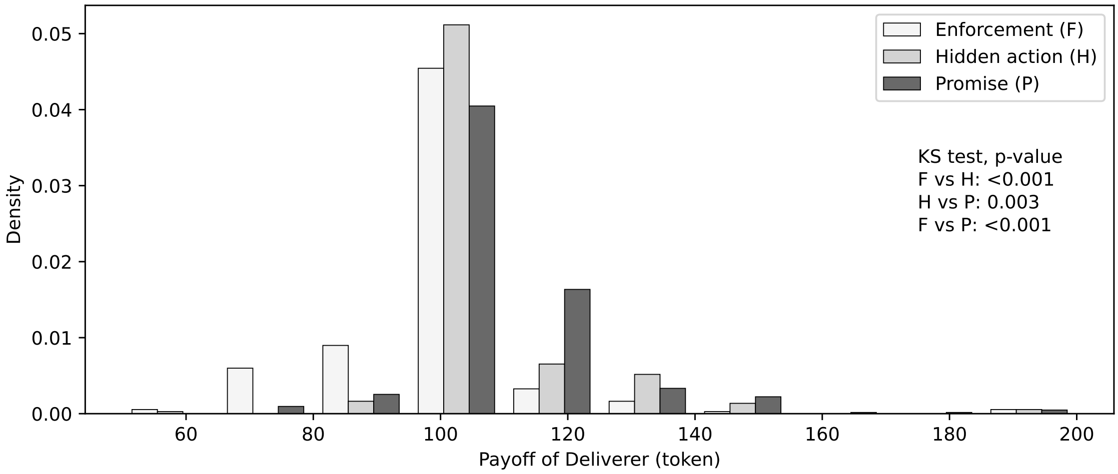

Figure 6: Distribution of Payoff of Deliverer. *Note*: We show the p-value of the two-sample Kolmogorov–Smirnov test in the upper right corner.

Table 8: Regressions of Allocation Ratio

	Allocation ratio (percent)		
	(1)	(2)	(3)
Hidden action	-74.551***	-74.960***	-74.494***
	(12.393)	(12.958)	(7.572)
Promise	-67.731***	-67.777***	-65.719***
	(16.750)	(16.934)	(13.589)
Amount sent		-0.095	-0.070
		(0.179)	(0.172)
Control average	113.743	113.743	113.743
F-test, p-value			
Hidden action = Promise	0.328	0.291	0.233
Covariates			X
Observations	910	910	910
Adjusted R^2	0.160	0.160	0.167

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered by sessions are in parentheses. Covariates are round fixed effects, patron's and dictator's academic fields (social science dummy, economics dummy, and natural science dummy).

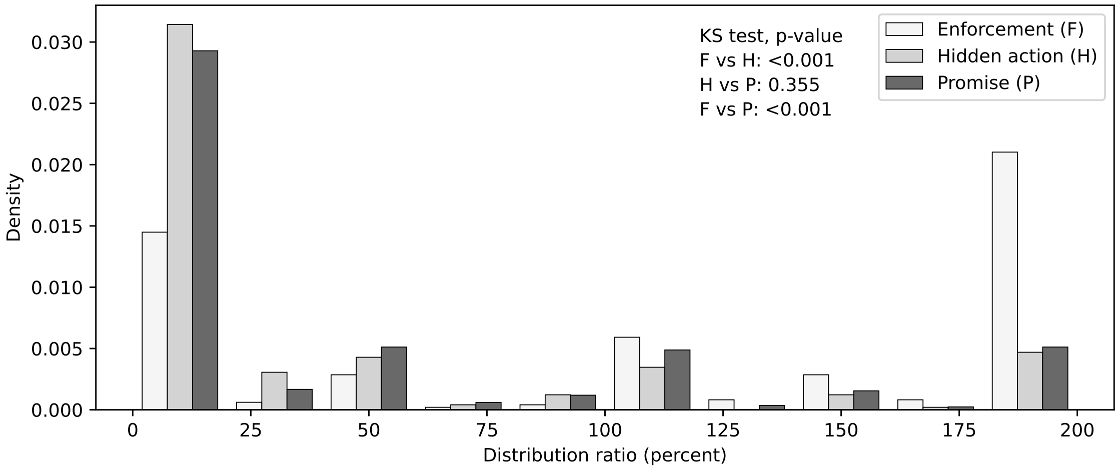

Figure 7: Distribution of Allocation Ratio. *Note*: We show the p-value of the two-sample Kolmogorov–Smirnov test in the upper right corner.

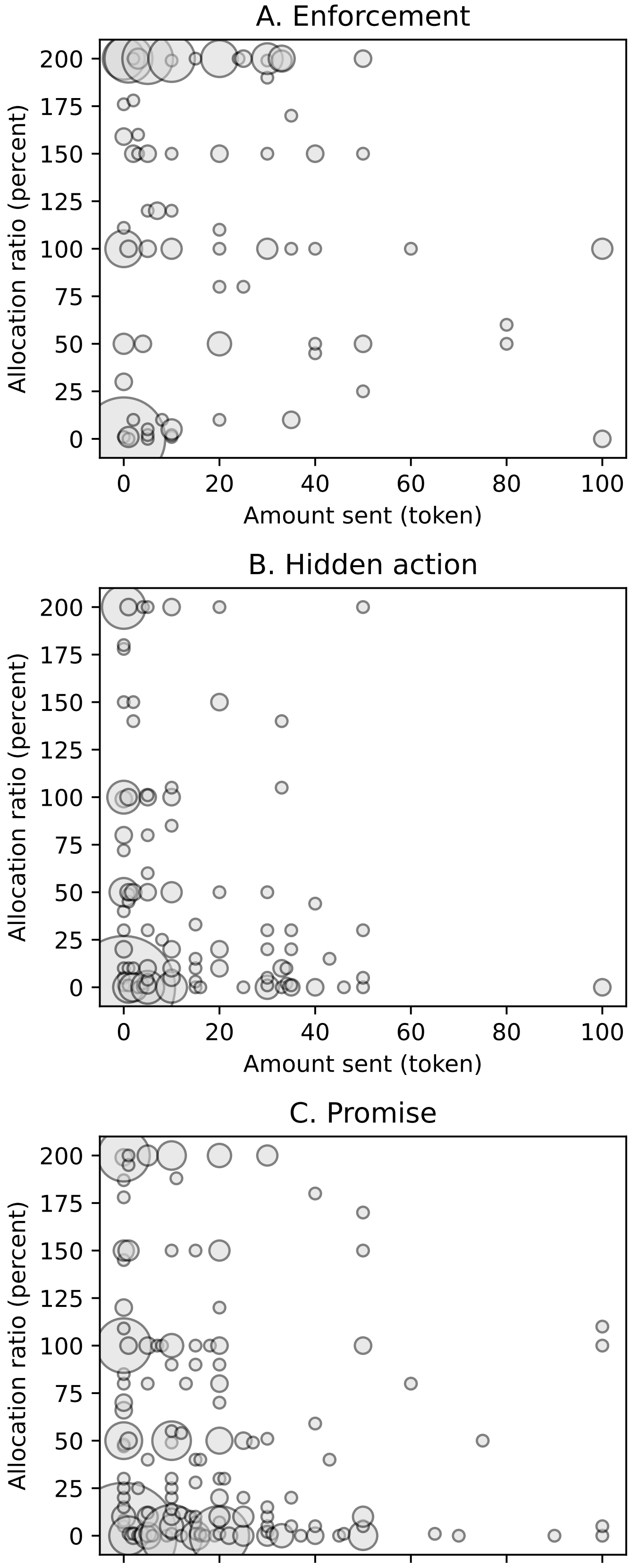

Figure 8: Joint Distribution of Sent Amount and Distribution Ratio by Each Treatment Group. *Note*: The bubble size is the sample size.

Table 9: Classification of Deliverer's Motivation

	Hidden action	Promise
No transaction	121	164
Equitable outcomes can be realized	18	34
Feasible: Selfish	8	13
Feasible: Inequality aversion	0	2
Feasible: Maximin criterion	2	3
Feasible: Other altruistic type	8	16
Equitable outcomes cannot be realized	106	222
Infeasible: Selfish	38	99
Infeasible: Possible inequality aversion or maximin criterion	7	17
Infeasible: Other altruistic type	61	106

Note: When $x = 0$, there are no transactions. We consider a deliverer who chooses $s = 0$ to be selfish. When $x > 1/3$, the deliverer can achieve an equal outcome at an interior point. When $x \leq 1/3$, the deliverer cannot achieve an equal outcome. In this case, a deliverer concerned with inequality aversion or the maximin criterion may choose $s = 2$ to minimize inequality as much as possible. We consider a deliverer who chooses $s = 2$ under $x \leq 1/3$ to be concerned with inequality aversion or the maximin criterion. Among deliverers who choose $s > 0$, we consider those who cannot be explained by inequality aversion or the maximin criterion to be other altruistic type.

B Theory of Dictator's Altruistic Behavior

Here, we theoretically demonstrate the possibility that altruistic dictators, who consider the payoff of the receiver, have an incentive to send transfers to the deliverer under $s > 0$. Therefore, we discuss disregarding the equilibrium of the game and dictator's expectations regarding s.

Pure altruism model and impact philanthropy model. The pure altruism model can be written as $U_D(\pi_D, \pi_R)$ for the dictator's utility function. In our experiments, the impact philanthropy model has a same utility function. This model explicitly incorporates the production function of public goods, treating donations as inputs. In our experiment, the production output of donations is the receiver's payoff, and the production function is linear, i.e., $Z(x) = s \cdot x$. This model suggests that dictators care about the increase in output due to contributions (impact philanthropist), proposing a utility function $U_D(1 - x, \delta)$, where $\delta = Z(x) - Z(0)$. This also aligns with the pure altruism model: $U_D(1 - x, sx)$.

Here, consider CES-type pure altruism model:

$$U_D(\pi_D, \pi_R) = [(1-\beta)\pi_D^{\rho} + \beta\pi_R^{\rho}]^{(1/\rho)}, \tag{1}$$

where $\beta \in (0, 1)$ is the altruistic parameter, and $\rho \in [-1, +\infty)\ \{0\}$ is the substitutability parameter. Then, the marginal utility of contribution is:

$$\frac{\partial U_D}{\partial x} = [-(1-\beta)\pi_D^{\rho-1} + s\beta\pi_R^{\rho-1}]U_D^{(1-\rho)/\rho}. \tag{2}$$

When $s = 0$, the marginal utility of x is always negative, so there is no incentive to choose $x > 0$. When $s > 0$, the marginal utility of x may be positive, so an incentive to choose $x > 0$ arises.

Inequality aversion. In this model, dictators make decisions according to the following utility function:

$$U_D = \begin{cases} \pi_D - \frac{\alpha}{2}[(\pi_E - \pi_D) + (\pi_R - \pi_D)] & \text{if} \quad x(1+s) > 1 \\ \pi_D - \frac{\alpha}{2}(\pi_E - \pi_D) - \frac{\beta}{2}(\pi_D - \pi_R) & \text{if} \quad x(1+s) \leq 1 \end{cases}. \tag{3}$$

Here, β represents the degree to which dictators dislike advantageous inequality, with $\beta \in [0, 1)$. Additionally, α represents the degree to which dictators dislike disadvantageous inequality, satisfying $\beta \leq \alpha$. For any x and s, $\pi_E - \pi_D = (2 - s)x \geq 0$. However, if $x(1 + s) \leq 1$, then $\pi_R - \pi_D = (s + 1)x - 1 \leq 0$.

When $s = 0$, $\pi_E - \pi_D = 2x$ and $\pi_D - \pi_R = 1 - x$. Thus, the increase in contributions widens the inequality between dictators and deliverers (disadvantageous inequality for dictators) while reducing the inequality between dictators and receivers (advantageous inequality for dictators). The negative marginal utility from the former inequality enlargement outweighs the positive marginal utility from the latter inequality reduction, so there is no incentive to choose $x > 0$ under $s = 0$. As the allocation ratio increases, the enlargement of inequality between dictators and deliverers due to increased contributions is mitigated (reduction in negative marginal utility), while the reduction of inequality between dictators and receivers is accelerated (increase in positive marginal utility). Therefore, an incentive to choose $x > 0$ arises under $s > 0$.

Social welfare. Consider the Charness-Rabin type of social preference model as follows:

$$U_E = \begin{cases} (1-\lambda)\pi_D + \lambda[\delta\pi_D + (1-\delta)\cdot 2] & \text{if} \quad s(1+x) > 1 \\ (1-\lambda)\pi_D + \lambda[\delta\pi_R + (1-\delta)\cdot 2] & \text{if} \quad s(1+x) \leq 1 \end{cases}. \tag{4}$$

Here, $\lambda \in [0, 1]$ represents the degree of importance given to social welfare, and $\delta \in (0, 1)$ represents the degree of importance given to the worst-off individual (maximin criterion). When dictators choose or anticipate $s = 0$, $\pi_R = 0$. Therefore, there is no incentive to choose $x > 0$. Under positive allocation ratios, in the region $s(1 + x) \leq 1$, an increase in contributions contributes to an increase in π_R, giving dictators an incentive to choose $x > 0$.

C Theory of Deliverer's Altruistic Behavior

The higher the amount sent by the dictator, the lower the allocation ratio the deliverer chooses. Here, we theoretically justify this result using several social preference models.

Let $s^* \in [0, 2]$ be optimal when $x = 0$, as the allocation ratio s does not affect player payoffs. Hence, we focus on the case where $x > 0$.

CES utility function. Assume $U_E(\pi_R)$ follows a CES-type utility function:

$$U_E(\pi_R) = [(1 - \beta)(1 + x - \pi_R)^\rho + \beta\pi_R^\rho]^{(1/\rho)}, \tag{5}$$

where $\pi_E = 1 + x - \pi_R$. Here, $\beta \in (0, 1)$ is the altruistic parameter, and $\rho \in [-1, +\infty)$ $\{0\}$ is the substitutability parameter. Given x, the optimal first-order condition for s is:

$$\frac{\partial U_E}{\partial \pi_R} = [-(1 - \beta)(1 + x - \pi_R)^{\rho-1} + \beta\pi_R^{\rho-1}]U_E^{(1-\rho)/\rho} = 0. \tag{6}$$

Thus, the optimal pie share ratio (ratio of pie $1 + x$ sent to the receiver) $p = \pi_R/(1 + x)$ is:

$$p^* = \left[1 + \left(\frac{\beta}{1 - \beta}\right)^{-1/(1-\rho)}\right]^{-1}, \tag{7}$$

or the optimal allocation ratio s^* is

$$s^* = \left[1 + \left(\frac{\beta}{1 - \beta}\right)^{-1/(1-\rho)}\right]^{-1} \frac{1 + x}{x}. \tag{8}$$

The optimal allocation ratio is a decreasing function of the dictator's transfer amount x:

$$\frac{ds^*}{dx} = -\frac{\alpha}{x^2} < 0, \tag{9}$$

where $\alpha = (1 + (\beta/(1 - \beta))^{-1/(1-\rho)})^{-1} > 0$.

Inequality aversion. Consider the Fehr-Schmidt type of inequality aversion model as follows:

$$U_E = \begin{cases} \pi_E - \frac{\alpha}{2}(\pi_R - \pi_E) - \frac{\beta}{2}(\pi_E - \pi_D) & \text{if} \quad s > \frac{1+x}{2x} \\ \pi_E - \frac{\beta}{2}[(\pi_E - \pi_D) + (\pi_E - \pi_R)] & \text{if} \quad s \leq \frac{1+x}{2x} \end{cases}. \tag{10}$$

Here, β represents the degree to which the deliverer dislikes advantageous inequality and satisfies $\beta \in [0, 1)$. The parameter α indicates the degree to which the deliverer dislikes disadvantageous inequality and satisfies $\beta \leq \alpha$. Given x, the marginal utility of s is:

$$\frac{\partial U_E}{\partial s} = \begin{cases} x\left(\frac{\beta}{2} - 1 - \alpha\right) < 0 & \text{if} \quad s > \frac{1+x}{2x} \\ x\left(\frac{3}{2}\beta - 1\right) & \text{if} \quad s \leq \frac{1+x}{2x} \end{cases}. \tag{11}$$

Therefore, if $2/3 < \beta$, the optimal allocation ratio is $s^* = (1 + x)/2x$, which is a decreasing function of x ($ds^*/dx = -1/2x^2 < 0$). If $\beta \leq 2/3$, the optimal allocation ratio is $s^* = 0$.

Social welfare. Consider the Charness-Rabin type of social preference model as follows:

$$U_E = \begin{cases} (1-\lambda)\pi_E + \lambda[\delta\pi_D + (1-\delta)\cdot 2] & \text{if} \quad s > \frac{1-x}{x} \\ (1-\lambda)\pi_E + \lambda[\delta\pi_R + (1-\delta)\cdot 2] & \text{if} \quad s \leq \frac{1-x}{x} \end{cases}. \tag{12}$$

Here, $\lambda \in [0, 1]$ represents the degree of importance given to social welfare, and $\delta \in (0, 1)$ represents the degree of importance given to the worst-off individual (maximin criterion). Given x, the marginal utility of s is:

$$\frac{\partial U_E}{\partial s} = \begin{cases} -x(1-\lambda) < 0 & \text{if} \quad s > \frac{1-x}{x} \\ x(\lambda\delta - (1-\lambda)) & \text{if} \quad s \leq \frac{1-x}{x} \end{cases}. \tag{13}$$

Therefore, if $(1 - \lambda)/\lambda < \delta$, the optimal allocation ratio is $s^* = (1 - x)/x$, which is a decreasing function of x ($ds^*/dx = -1/x^2 < 0$). If $\delta \leq (1 - \lambda)/\lambda$, the optimal allocation ratio is $s^* = 0$.