



UNIVERSITY OF GOTHENBURG  
SCHOOL OF BUSINESS, ECONOMICS AND LAW

**WORKING PAPERS IN ECONOMICS**

**No 664**

**Why (field) experiments on unethical behavior are important: Comparing stated and revealed behavior**

**Yonas Alem, Håkan Eggert, Martin G. Kocher, and Remidius D. Ruhinduka**

**June 2016**

**ISSN 1403-2473 (print)**  
**ISSN 1403-2465 (online)**

# Why (field) experiments on unethical behavior are important: Comparing stated and revealed behavior\*

Yonas Alem<sup>a</sup>, Håkan Eggert<sup>b</sup>, Martin G. Kocher<sup>c</sup>, and Remidius D. Ruhinduka<sup>d</sup>

June 21, 2016

**Abstract:** Understanding unethical behavior is essential to many phenomena in the real world. We carry out a field experiment in a unique setting that varies the levels of reciprocity and guilt in an ethical decision. A survey more than one year before the field experiment allows us to compare at the individual level *stated* unethical behavior with *revealed* behavior in the same situation in the field. Our results indicate a strong discrepancy between stated and revealed behavior, regardless of the specific treatment in the field experiment. This suggests that, given a natural setting, people may actually behave inconsistently with the way in which they otherwise “brand” themselves. Our findings raise caution about the interpretation of stated behavioral measures commonly used in research on unethical behavior. However, we show that inducing reciprocity and guilt leads to a decrease in unethical behavior.

**JEL Classification:** C93, D01, D03

**Keywords:** Honesty; kindness; guilt; field experiment; behavioral economics

---

<sup>a</sup> Department of Economics, University of Gothenburg, Sweden, e-mail: [yonas.alem@economics.gu.se](mailto:yonas.alem@economics.gu.se).

<sup>b</sup> Department of Economics, University of Gothenburg, Sweden, e-mail: [hakan.eggert@economics.gu.se](mailto:hakan.eggert@economics.gu.se).

<sup>c</sup> Department of Economics, University of Munich, Germany, e-mail: [martin.kocher@lrz.uni-muenchen.de](mailto:martin.kocher@lrz.uni-muenchen.de); Department of Economics, University of Gothenburg, Sweden; School of Economics and Finance, Queensland University of Technology, Brisbane, Australia.

<sup>d</sup> Department of Economics, University of Gothenburg, Sweden, Department of Economics, University of Dar es Salaam, Tanzania. e-mail: [remidius@yahoo.com](mailto:remidius@yahoo.com).

\* Many thanks to Martin Dufwenberg, Glenn Harrison, Michel Marechal, Haileselassie Medhin, Travis Lybbert, Matthias Sutter, and seminar participants at the University of Gothenburg, the University of Munich, and the “Morality, Incentives and Unethical Behavior Conference” at the University of California, San Diego, for valuable comments. Financial support from the Swedish International Development Agency (SIDA) through the Environmental Economics Unit of the Department of Economics, University of Gothenburg, and from Formas, through the program Human Cooperation to Manage Natural Resources (COMMONS), is gratefully acknowledged.

## 1. Introduction

There are many decisions in which agents have to choose between behaving honestly or following an ethics code and being dishonest. Unethical or dishonest behavior can come, for instance, in the form of lying, cheating, or pursuing one's own self-interest instead of following a focal social convention or norm. Economists are mainly interested in situations in which dishonest behavior provides higher monetary payoffs than honest behavior. In the tradition of Becker (1968), the usual assumption is that humans engage in unethical acts if those acts maximize monetary payoffs without negative consequences. However, it is straightforward to extend the individual utility function by a term that introduces (psychological) costs of lying or cheating. Then, individuals trade-off the benefits of an unethical act and its costs, and they should refrain from profitable acts of cheating if total costs are higher than benefits (Gneezy, 2005; Ariely, 2012; Abeler et al., 2014; Rosenbaum et al., 2014; Kajackaite and Gneezy, 2016).

There is a quickly growing literature in economics studying the determinants of the trade-off between ethical and unethical behavior in various domains. We know that there is considerable individual heterogeneity in the inclination to lie or cheat; and circumstances, framing, the monetary consequences of the trade-off, beliefs about the norm, peer behavior, and many other aspects matter in relevant decisions. The vast majority of the relevant studies rely on laboratory experiments (for early examples, see Gneezy, 2005; Sutter, 2009; Houser et al., 2012; Fischbacher and Föllmi-Heusi, 2013). The problem with naturally occurring data in the context of unethical behavior is that dishonest behavior or cheating (such as tax evasion, fraud, cheating on one's spouse, or lying about a signal that is important to another person) often cannot be observed or can only be observed partially, creating all sorts of problems with the interpretation of data. Randomized controlled trials (RCTs) in the field are a potential remedy, but so far they have been very scarce when it comes to studying dishonest behavior. Among the few recent exceptions in economics are Shu et al. (2012), Azar et al. (2013), and Pruckner and Sausgruber (2013), which are discussed in greater detail below.

Our study has three objectives. First, we have the unique opportunity to link revealed behavior in a moral dilemma in our RCT with stated behavior in the same dilemma based on a survey conducted more than one year before the field experiment. This link can be made at the individual level, and it allows us to assess the correlation between stated behavior and revealed behavior in a moral context. As far as we know, we are the first to do so, and our

specific setup (the stated question embedded in a much larger survey and the long time between eliciting stated behavior and revealed behavior) makes it very unlikely that participants draw a connection between the survey and the experiment. This excludes concerns regarding a potential preference for consistent behavior across the two elicitation methods. Second, we assess the impact of different “frames” (and some additional incentives) on the inclination to behave ethically. Our RCT provides three frames: a neutral frame, a kindness frame, and a frame that we call “guilt”. We are able to analyze which of these frames elicits the most honest behavior. Third, we want to add to the scarce existing evidence on unethical or dishonest behavior in the field, based on an RCT. We use a very natural setup to assess honesty and moral/ethical behavior in the field, and participants do not know that they are part of an experiment.

Our data captures a set of misdirected mobile phone payments in Tanzania. More precisely, we analyze the reaction of a sample of Tanzanians to receiving a money transfer that is obviously misdirected and to an SMS message that was sent immediately after the transfer asking for the return of the money. The SMS message was (i) framed neutrally, (ii) involved a gift (kindness), or (iii) tried to induce a feeling of guilt on the part of the recipient. Return rates and the amount of money returned are our measures for (dis)honest behavior. We have stated return inclinations from a survey and revealed behavior in the field experiment from the same people.

In the following, we discuss the threefold contribution of our paper. Our first contribution is related to the association between stated and revealed behavior in a moral context. Understanding this association is very relevant for the researcher. If answers to appropriate survey questions are highly correlated with revealed behavior, the former would be clearly more desirable than the latter: non-incentivized surveys are usually cheaper and less intrusive, and they usually imply fewer ethical concerns. However, if the correlation is weak (see, e.g., Bertrand and Mullainathan, 2001), it seems warranted to incur the various costs of conducting field experiments to gain more knowledge on moral behavior. There is a related discussion on the external validity of laboratory experiments in the field (e.g., Levitt and List, 2007; Falk and Heckman, 2009). Our focus here is not on the stability of behavior when going from the

laboratory to the field (lab-field)<sup>1</sup>, but on the comparison between two situations in the field (field-field) with different incentives.

The association between stated and revealed behavior is obviously a relevant issue in any research that relies on surveys. One example is the long and ongoing debate regarding the usefulness of stated preference methods in general and contingent valuation studies in particular for assessing non-use values when valuing the environment (Diamond and Hausman, 1994; Hanemann, 1994; Hausman, 2012; Kling et al., 2012). Some of the findings of the current study might thus be relevant beyond the realm of moral decision-making.

Our second contribution, related to the framing of the decision-making situation, has to do with investigating whether a kind act or the induction of guilt can reduce the inclination to behave unethically, compared to a baseline (control) condition. We draw on a large literature on the framing of donation decisions (Falk, 2007; List, 2011; Alpízar and Martinsson, 2012; Kube et al., 2012) and want to see whether some of the results from this literature also apply to a setup, which is not about giving, but about keeping something that does not belong to oneself. After introducing the general setup, we explain the details of our framing variation in Section 3.

As mentioned, the economics literature using RCTs to study unethical behavior in the field is limited.<sup>2</sup> Our third contribution is adding to the existing evidence on unethical behavior in a field setting. Haan and Kooreman (2002) study the likelihood of paying for candy bars by company employees in an honor system and find that a large proportion of employees do not pay and that their average payments decline over time. Levitt (2006) uses a similar setup to investigate honesty in paying for bagels and donuts by corporate employees. He observes that average payment for bagels declines with a rise in the price and increases with the fraction of uneaten donuts and bagels. More recently, Shu et al. (2012) analyze moral pledges when signing forms and show that signing a form with a proof of honest intent at the beginning

---

<sup>1</sup> See, for instance, Kröll and Rustagi (2015) or Dai et al. (2016) for recent contributions in a general context that is similar to ours.

<sup>2</sup> Obviously, the set of existing papers depends on the breadth of the definition of unethical and ethical behavior. Here, we do not discuss studies that look at charitable giving and allocation decisions, because all existing field experiments are concerned with donations (giving money) rather than, as in our case, dishonesty in keeping somebody else's money. There is also a literature on credence goods that is related to our research. Balafoutas et al. (2013) study the behavior of cab drivers in terms of fare dodging or taking detours when the passenger signals that he or she is a foreigner.

rather than at the end of the form increases honest reporting. Pruckner and Sausgruber (2013) look at stealing newspapers from dispatch units on the street. They also show that reminders regarding morality increase the level of honesty, while recaps of the legal norm have no significant effect on behavior. Azar et al. (2013) implement a field experiment that, in spirit, is close to our setup, even though our focus is very different. They let restaurant waiters intentionally return extra change to customers and show that those who receive 10 extra Shekels (about USD 3) are less likely to return them to the waiter than those who receive 40 extra Shekels (about USD 12).<sup>3</sup> In Abeler et al. (2014), the dice-throw paradigm (see, Fischbacher and Föllmi-Heusi, 2013) is used to study lying in a representative sample using telephone interviews. They observe surprisingly little evidence for lying among the respondents on the telephone, but the unusual request to throw dice and report the result to a stranger on the phone might have contributed to this result, despite the monetary incentives to report high numbers.<sup>4</sup>

Hanna and Wang (2014) provide interesting results that is based on a laboratory experiment, but where laboratory behavior is linked to decision-making outside the laboratory. They use the dice-throwing task to predict whether dishonest students are more likely to indicate that they want to work in the public sector, which is indeed the case. They also find that cheating in the dice-throwing task among public workers is associated with more corrupt behavior of these officials. In a somewhat similar vein, Franzen and Pointner (2014) look at the correlation between dictator giving behavior in the laboratory and honesty in returning an intentionally misdirected letter containing money. Indeed, dictator giving is associated with more honesty. Also, Kröll and Rustagi (2015) investigate the relationship between honest behavior in a laboratory task and outside the laboratory. They find that dishonest milkmen (measured by a version of the dice-throwing experiment) cheat more than honest milkmen by adding much more water to the milk. Finally, Potters and Stoop (2016) analyze the correlation between cheating in the laboratory and in the field by combining a laboratory decision, which allows detection of cheating, with an e-mail sent to participants regarding payment, which

---

<sup>3</sup> In a somewhat less controlled fashion, Reader's Digest magazine "lost" twelve wallets in each of 16 cities around the world and checked how many of them were actually returned to the owner (<http://www.rd.com/slideshows/most-honest-cities-lost-wallet-test/view-all/>). See Dufwenberg and Gneezy (2000) for a laboratory implementation of the situation.

<sup>4</sup> Bucciol et al. (2013) look at free-riding in public transportation in Italy. In their sample, 43% of passengers do not pay. However, the motivation to cheat in situations that involve a company or a public organization might be different than in a bilateral interaction among ex-ante equal individuals.

also allows measurement of cheating. Cheating in the lab and in the field are correlated at the individual level in their study.

Our empirical results indicate that there is a weak association between stated and revealed behavior. Thus, relying on stated behavior when drawing policy conclusions is far from optimal. Interestingly, people deviate from their stated behavior in two directions – they behave honestly even though they stated earlier that they would be dishonest, and they behave dishonestly even though they stated that they would be honest. However, the bulk of our observations go in the expected direction of stated honesty and revealed dishonesty. Our survey allowed the response “not sure what to do,” which was chosen by less than 10% of the respondents, but this group turned out to be by far the most honest in the RCT. With regard to the framing, we show that the actual levels of honesty or dishonesty are easily malleable. The specific effects depend on the framing that decision makers face, but we think that the general result of malleability, which also has been shown in laboratory experiments and in the few existing field experiments in economics on (dis)honest behavior, carries over to our setting. Socio-economic background variables and individual uncertainty preferences are weak predictors of ethical or unethical behavior in our data.

The rest of this paper is organized as follows. Section 2 describes the details of our field setup and our empirical identification assumptions. In Section 3, the design of our RCT is discussed, and Section 4 provides our empirical results. Finally, Section 5 discusses our findings and concludes the paper.

## **2. Measuring honesty in the field**

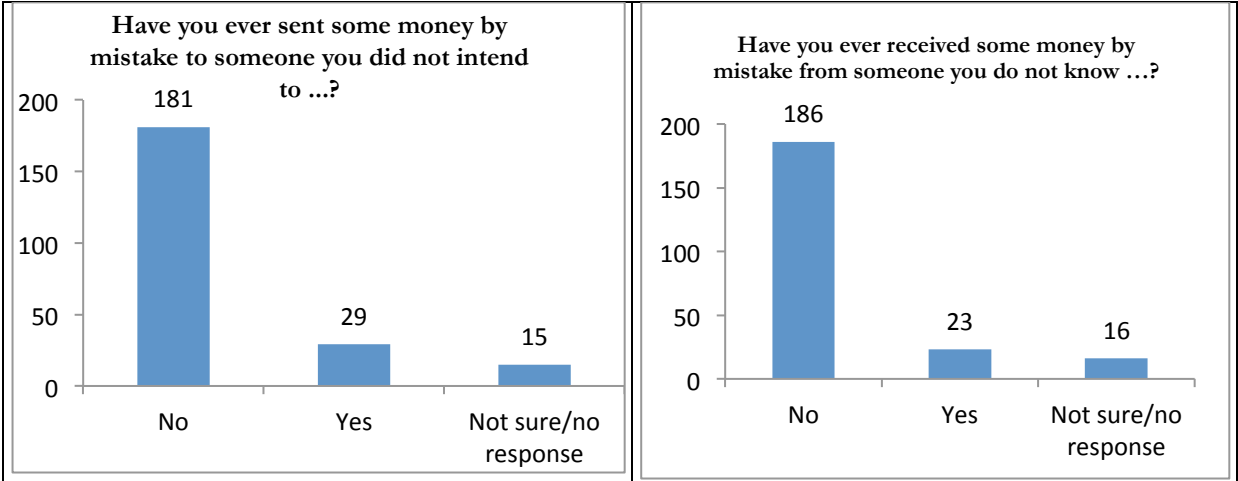
Both the stated and revealed behavior we investigate in this paper are related to unintended or potentially misdirected monetary transfers using mobile phone banking in Tanzania. Our variable of interest is the individual inclination to return an unintended transfer to the sender, despite the fact that the sender cannot enforce such a return. Hence, the return is voluntary, but we will use different one-time messages to induce returning the money. We will call returning the money “ethical behavior” and keeping the money (despite a request to return it) “unethical”, although this classification might not be fully appropriate in every single case.

Because we draw heavily on the mobile phone banking system in Tanzania, we first describe how it works. The use of mobile phone banking in Tanzania has grown rapidly since its first

introduction in 2008. According to the central Bank of Tanzania, by December 2013, the country had more than 30 million registered mobile phone bank accounts, of which more than 11 million were recorded as active (BoT, 2014). A recent survey by InterMedia (2013) asserts that 65% of households owned at least one account as of January 2013. Importantly, within only one month (December 2013), mobile money transfer businesses in the country performed transactions worth more than Tanzanian Shillings (TZS) 3 trillion (equaling USD 1.8 billion) (Di Castri and Gidvani, 2014).

Because of the large customer base and number of transactions in a day, unintended transfers are common. It is easy to make a mistake and transfer the money to person A instead of person B. The sender may then immediately learn about the mistake when (s)he receives a text message from the service provider confirming the transaction to a “wrong” number (or name). When this happens, the sender will have to ask the recipient to send back the money. The recipient then decides whether to send the money back or just keep it. In contrast to bank transfers, a mobile phone transfer cannot simply be recalled—a fact that both the sender and the receiver are well aware of.

**Figure 1: Past experience with unintended mobile phone based money transfers**



Source: Authors’ construction based on survey data from <blinded for anonymity>.



Figure 1 presents data from two survey questions<sup>5</sup>, asking whether respondents have either sent or received money by mistake through their mobile phone accounts within the past year. It is obvious that incorrect transfers are common. More than 50 out of 225 respondents have either sent money unintentionally or received money from a source that had sent it unintentionally, and three had experienced both within a year.

While not returning money in such circumstances is a crime in countries such as the UK and the US, the legal obligation is vague in Tanzania. The situation is even worse in the mobile phone banking system. For example, Airtel Telephone Company states in its terms of service that any amount of “*airtel-money*” transferred erroneously by the customer shall remain the sole responsibility of the customer and the company will have no liability whatsoever regarding the transaction.<sup>6</sup>

Our identifying assumption stipulates that whether or not the recipient sends back the money is mainly an ethical question. One potential concern is whether what we capture is indeed a measure of dishonesty. People could be risk averse, i.e., worried that they may be caught and legally charged for theft should they decide to keep the money. In response, it is important to note that the mobile bank system in Tanzania was introduced prior to the establishment of a regulatory framework by the central bank to govern its operations (Di Castri and Gidvani, 2014). Upon the initial proposal by Vodacom Telephone Company, the central bank issued a so-called “letter of no objection” to the company and voiced its intention that a legal framework would be formulated later. To date, there exists no legal framework governing mobile banking transactions, and the recipient can hardly be charged with legal offenses in such circumstances. Also, as mentioned, mobile phone companies articulate that any amount transferred erroneously by the customer shall remain the sole responsibility of the customer. It is thus generally perceived that the probability of being caught and legally prosecuted is close to zero.

A related concern, even in the absence of formal punishment, is that the sender could privately look for the recipient and punish that person. However, in the mobile phone framework in Tanzania, the search costs to the sender are too high to pursue such a motive,

---

<sup>5</sup> Data are based on a multi-purpose survey, aimed at assessing the adoption and impact of a climate-friendly agricultural technology, called the System of Rice Intensification (SRI). Details are provided in Section 3.

<sup>6</sup> See <http://www.airtel.in/personal/money/terms-of-use> and <http://africa.airtel.com/wps/wcm/connect/africarevamp/Tanzania/AirtelMoney/faq/>

especially in our case where we only transferred about 20,000 TZS (USD 12). It is very difficult to obtain information on the whereabouts of the recipient, although mobile phone accounts are registered. This is common knowledge to both parties. Given the setup, it is not surprising that a good share of mistakenly transferred money never gets returned to the sender. Further, in case risk or ambiguity preferences do nonetheless matter for the decision to return money, our data also allow us to directly control for their influence, because they were elicited in the survey (in an incentivized way).

### **3. Participants, experimental design, and hypotheses**

#### *3.1 Participants*

We secured clearance from the Research Board at the University of Dar es Salaam and conducted our experiment with heads of farming households from rural Tanzania (specifically from eight villages of the Morogoro region). Our participants took part in a large, multi-purpose survey, involving a monetarily incentivized elicitation of risk, ambiguity and time preferences in September 2013. This initial survey involved 338 randomly selected households. During the survey, each participant was asked whether s/he owns an active mobile phone account that is registered for mobile phone banking. About 90 percent (302 participants) provided an affirmative answer, and the remaining household heads indicated that they usually use either relatives' or neighbors' phones to make calls or carry out such transactions. The mobile phone numbers used by all participants were recorded in the survey, regardless of who owned the phone. This high share of access to the system motivated the use of mobile phone banking to deliver delayed payments for a time preference elicitation experiment in 2013.

In order to study unethical behavior, we ran the experiment, including only those respondents who reported privately owning a mobile phone account, while dropping those who used another person's account. Three days prior to the start of the experiment, we conducted a pre-test of the recorded numbers to check whether they were still active and registered for mobile phone banking. We did it twice on that day (November, 3<sup>rd</sup>, 2014), calling from different phone numbers, which were not used in the actual experiment. From this pre-test, 226 phone

numbers were found to be still active and registered for mobile phone banking.<sup>7</sup> We then randomly assigned 225 farmers into three treatment groups.

### 3.2 *Experimental design*

To conduct our experiment, we bought and registered nine different mobile phone sim-cards, three from each of the three major mobile phone companies providing service in the area (Vodacom, Airtel, and Tigo). In order to avoid any potential bias due to the sex of the sender, all numbers were registered with male names, given that 93% of the participating household heads in the experiment are male. None of the sim-cards had been used during the payment for the survey and preference elicitation in 2013. We then topped up the accounts with money and sent the money to the farmers. Each received exactly the same amount of money, TZS 20,000 (equivalent to about USD 12)<sup>8</sup>. Once the money is sent, the network system sends back a confirmation message that a specified sum of money has been sent to the owner of a particular mobile phone. We then immediately sent a text message to the receiver asking them to return the money (less TZS 500 covering the transaction cost). Our three “framing” treatments differ in the message sent (see Table 1). We used sim-cards registered under the same service provider as that of a given participant/recipient. This enabled us to confirm the names of our participants on their accounts before we made the transfer; hence, we could be sure that we were sending the money to the right person. Delivery status of the sent messages provided an extra confirmation of the transfer.

Our three treatments vary the message to the receiver of the transfer. The CONTROL treatment implements a friendly but rather neutral message to the receiver. It reads “Hi, I have just transferred TZS 20,000 to your m-pesa (or tigo-pesa, etc.) account. I was not supposed to send it to you. Could you kindly transfer back TZS 19,500 and use TZS 500 for the transfer fee? Thank you very much.” Our KINDNESS treatment intends to invoke reciprocity by giving a “gift” to the recipient. The message is very similar (see Table 1), but we offer the recipient to keep TZS 5,000 (plus the fee of TZS 500) and only return TZS 14,500 to the

---

<sup>7</sup> During the experiment and even two days later, the money and text message could not be delivered to one of the 226 subjects. We dropped the observation, leaving us with 225 participants.

<sup>8</sup> The average household expenditure on daily basic needs in our study area was TZS 4,500 at the time of the survey. Hence, the transferred amount corresponds to more than four days of expenditures for an average household.

sender. We chose 25% as a share to be retained, because it seemed a good compromise between being large enough to really matter for the recipient and still small enough to make it worthwhile for the sender to make the offer (see also Fehr and Gächter, 2000). The third treatment, GUILT, aims to induce a stronger feeling of guilt in the recipient. We alter the message by stating that the money was intended for one of the largest orphanage centers in Tanzania, the *Msimbazi Orphanage Centre*, to support poor children. The total amount of money returned by the subjects in this treatment later on was indeed donated to the Msimbazi Orphanage Centre in Dar es Salaam, Tanzania.

**Table 1: Treatments in the field experiment: English translation of the sent messages**

| <b>Treatment</b> | <b>Message sent</b>   |
|------------------|---|
| CONTROL          | <i>Hi, I have just transferred TZS. 20,000 to your m-pesa (or tigo-pesa, etc) account. I was not supposed to send it to you. Could you kindly transfer back TZS. 19,500 and use TZS.500 for the transfer fee? Thank you very much.</i>  |
| KINDNESS         | <i>Hi, I have just transferred TZS. 20,000 to your m-pesa (or tigo-pesa, etc) account. I was not supposed to send it to you. Could you kindly transfer at least 14,500 back to me and you may keep TZS 5000 as my token of appreciation? Thank you very much.</i>   |
| GUILT            | <i>Hi, I have just transferred TZS. 20,000 to your m-pesa (or tigo-pesa, etc) account. I was not supposed to send it to you but rather to the head of Msimbazi Orphanage Centre to help those poor orphan children. Could you kindly transfer back TZS. 19,500 and use TZS.500 for the transfer fee? Thank you very much.</i> |

The experiment was conducted within 20 hours divided into two days starting from the afternoon of November 6<sup>th</sup> to the morning of November 7<sup>th</sup>, 2014. In order to minimize spillover of the information across participants within the same village, we made sure, to the extent possible, that all the participants within the same village were sent the money on the same day and at the same hour<sup>9</sup>. Although mistaken transactions are common, behavior of experimental participants could be biased if they knew in advance of another person who

<sup>9</sup> We tested whether the day on which the experiment was conducted matters for the probability of returning the money. The results presented in Table 6 below show no influence of the date.

received the same amount of money and a similar message. Participants were never contacted again, and telephone numbers were deleted after a couple of weeks.

In the survey conducted back in September 2013, we asked all participants a hypothetical question on what they would do in case they received TZS 100,000 (about USD 59 at the time of the survey) in their account by mistake. The answer to this question is used in the comparison between stated and revealed behavior. The survey also allows us to control for background variables such as income, education, religious attitudes, and uncertainty preferences.

We decided to actually send “only” TZS 20,000 instead of TZS 100,000, after realizing from the results in the survey that TZS 20,000 is already equivalent to about four days of household expenditure. The larger amount could have created too much attention in the villages, with people talking about it, and it also would have induced a much more severe moral dilemma for the participants, which we wanted to avoid for research ethics reasons.

### 3.3 Predictions

In line with the existing literature, we expect some returns in CONTROL. Notice that ethical behavior in our cases requires an active decision, whereas doing nothing implies unethical behavior. This is in contrast to many other studies in which a decision has to be taken, and doing nothing is not an option. Hence, *a priori* it is difficult to predict the inclination to return the money. With respect to the KINDNESS treatment, it is also not clear what behavioral response the gift is likely to induce. It could increase the propensity to return the money (see, for instance, Falk, 2007; Kube et al., 2012) unless the monetary gift crowds out an intrinsic inclination to return (i.e., crowds out the potential warm glow from returning the money; see, e.g., Mellström and Johannesson, 2010, with respect to blood donations). Return levels could even be lower than in CONTROL due to self-image concerns, but our working hypothesis is that recipients are more likely to return in KINDNESS than in CONTROL.

Some recent laboratory experiments conducted in developed countries suggest that people’s revealed level of honesty may be affected when they learn that a third party is likely to be affected by their behavior, either positively or negatively (e.g., Gino and Pierce, 2009; Wiltermuth, 2011; Gino et al., 2013; Kajackaite and Gneezy, 2016). Other studies have revealed that induced guilt can significantly affect how a person may behave toward others

(e.g., Cunningham et al, 1980; Rebege et al, 2013). Whether our treatment variation really induces guilt or some similar feelings is a question on its own. For ease of reference, we refer to the treatment as GUILT, but it may well be that other aspects play a role. We expect return rates to be higher in GUILT than in CONTROL, assuming that the treatment indeed induced feelings associated with guilt or compassion among the recipients. We remain agnostic with regard to the relationship between the return rates in KINDNESS and GUILT.

*Hypothesis 1: In contrast to selfish predictions, there is a positive return rate in CONTROL.*

*Hypothesis 2: A “gift” in the form of an offer to share money in KINDNESS increases the return rate significantly over the one in CONTROL.*

*Hypothesis 3: Inducing guilt in the recipient by stating a good cause for which the money was intended increases the return rate in GUILT significantly over the one in CONTROL.*

Now we turn to our research question concerning the relationship between stated and revealed behavior. It is difficult to predict the relationship between stated and revealed behavior when it comes to moral issues. We are not aware of a previous study that addresses this aspect in a comparable way. Assuming that surveys can elicit truthful answers even in sensitive contexts, we expect that there is a correlation between stated and revealed behavior at the individual level, but that the correlation is far from perfect. We can also speculate about the specific form of the correlation. It seems reasonable to assume that those who indicate in the survey that they will be dishonest will indeed behave dishonestly. For those who indicate that they will behave honestly and return the money, we can expect to see some to follow their stated behavior and some to deviate (potentially because they stated the socially desirable behavior but are not willing to follow it when they are put in the real situation).

*Hypothesis 4: Stated and revealed behaviors in all the three treatments are correlated at the individual level. Stating being dishonest is expected to be strongly correlated with revealed behavior, while stating being honest is expected to be weakly correlated with revealed behavior.*

### 3.4 *Research ethics aspects*

Laboratory and field experiments on cheating and moral behavior create a trade-off from the perspective of research ethics. They naturally put decision makers into situations that involve a moral conflict. It is particularly this conflict that one wants to study, and we argue that the moral conflict is significant, but not as severe as to generate serious psychological discomfort in our case. Note that any resource allocation experiment (such as the dictator game) involves a similar conflict.

In our case, we link data from a questionnaire with decisions after the receipt of the money. To address our research question summarized in Hypothesis 4, it is necessary to make this link at the individual level without the consent of participants. The latter is the case in many field experiments in economics. The potential psychological effect of the study on participants seems acceptable because it is expected to be very limited. All safeguards regarding anonymity (removal of mobile phone numbers and names from the data after the completion of the experiment) have been implemented.<sup>10</sup>

All experiments on unethical behavior inside and outside the laboratory also involve an intentionally misleading signal regarding the real intention of the experimenters. The dice-throw paradigm signals that researchers are interested in dice throws, but they are actually interested in cheating. Losing one's wallet to study return behavior signals that the loss was by mistake, but actually it was intentional. The same is true for our transfers (and the messages). Notice that as mentioned in the preceding section, the money stated to be for the orphanage and returned by the recipients in treatment GUILT was actually transferred to the orphanage after the experiment.

## 4. **Experimental results**

We organize the results from our study along the following lines. In Section 4.1, we present an overview of descriptive variables on the return levels in order to investigate Hypotheses 1-3. In Section 4.2, we empirically link stated behavior in the survey with revealed behavior in

---

<sup>10</sup> The University of Dar es Salaam's Research Board (a body similar to IRBs) approved the experiment.

the field experiment. Finally, in Section 4.3, we take another look at the determinants of unethical behavior.

#### *4.1 Descriptive overview*

We have 13 out of 225 cases in which we are not sure whether the person received the money. For the descriptive overview, we proceed on the assumption that those 13 people indeed got the transfer. Later on, we will drop them for robustness checks, but we will always indicate clearly when we do this.

Our random allocation to the three treatments seems to have been successful. Out of 24 comparisons, only one – household daily spending – was significantly different between the CONTROL and GUILT groups (see Table A.1 in the Appendix, where means and standard deviations for relevant variables are displayed).

When we look at return levels, we note that most people who returned money returned the requested amount. In CONTROL, a total of 18 participants returned money, of whom 10 returned the requested amount of TZS 19,500, five returned TZS 19,000 and three returned TZS 20,000. In KINDNESS, we observe more variation, with returns between TZS 14,500 and 20,000 among those who returned money. In the GUILT treatment, 24 participants returned TZS 19,500, four returned TZS 20,000 and the remaining 47 returned nothing. Given the lack of variation in individual return levels when a positive amount is returned, which is deliberately induced by our setup, we consider return a binary variable. Both KINDNESS and GUILT result in higher return rates and total amount returned compared to CONTROL. Return rates were highest in the KINDNESS treatment, but given that many of those returning the money in the KINDNESS treatment kept the ‘gift’ of TZS 5,000, the total amount returned was highest from the GUILT treatment.

Table 2 provides an overview of the averages in the three treatments and the results of two-sided significance tests. In general, we find that 34.7 % of our sample returned some money to the sender. Only 24% of the participants in CONTROL returned the requested amount (or a very similar amount). Consistent with Hypothesis 1, this is clearly above zero – it is more than just behavioral noise – although the great majority did not return the transfer.



*Result 1: Consistent with Hypothesis 1, some recipients in the CONTROL treatment returned a positive amount of money to the sender, although a majority of them do not return anything.*

KINDNESS induces a higher return rate, consistent with Hypothesis 2. The rate almost doubles to 42.7% compared to CONTROL, and the difference between the two treatments is highly statistically significant. Offering a “gift” induced reciprocity among recipients. In addition, the mean amount returned is greater in KINDNESS, although the difference is not significant at conventional levels. The mean returned amount is TZS 6,661 in KINDNESS, compared to a mean returned amount of TZS 4,667 in CONTROL. Despite the lack of significance, the benefit of the “gift” obviously outweighs its costs on average.

*Result 2: KINDNESS induces significantly higher return rates than CONTROL, and the sender is better off on average in KINDNESS. This is despite the lower absolute amounts returned as a consequence of the “gift”.*

**Table 2: Distribution of return rate by treatment**

| <b>Treatment</b> | <b>Number of observations</b> | <b>Returned</b> | <b>Probability of returning</b> | <b>p-value of FE-test</b> | <b>Returned amount</b> | <b>p-value of MWU-test</b> |
|------------------|-------------------------------|-----------------|---------------------------------|---------------------------|------------------------|----------------------------|
| CONTROL          | 75                            | 18              | 24,0%                           | -                         | 4,667                  | -                          |
| KINDNESS         | 75                            | 32              | 42,7%                           | 0.02**                    | 6,661                  | 0.23                       |
| GUILT            | 75                            | 28              | 37,3%                           | 0.11                      | 7,567                  | 0.03**                     |
| All              | 225                           | 78              | 34,7%                           |                           |                        |                            |

Note: The p-values refer to two-sided Fischer exact (FE) and Mann-Whitney-U (MWU) tests for the differences in proportions and in means (medians) between the control group and the treatment groups.

\*\* significant at the 1% level.

Treatment GUILT lies in between. It induces a return rate of 37.3%, which is higher than the rate in CONTROL, as proposed in Hypothesis 3, but the difference between GUILT and CONTROL misses conventional levels of significance. On the other hand, GUILT creates the highest average amount returned among the three treatments, at TZS 7,567, which is significantly higher than in CONTROL and is consistent with Hypothesis 3.

*Result 3: GUILT induces higher return rates than CONTROL. It shows the highest average returned amounts of all three treatments.*

The general impression is that both treatments work in the same direction. From the perspective of the return rate, treatment KINDNESS works best, whereas from the perspective of the amount returned, treatment GUILT works better, because the cost of the gift cannot be compensated by the higher return rate in KINDNESS, when compared to GUILT.

#### *4.2 Do people act as they brand themselves? Stated versus revealed honesty*

Table 3 provides a comparison of the fractions from the stated honesty and from the revealed honesty. We pool data from all treatments for Table 3, but qualitatively there is not much difference. However, we assume that sending back TZS 14,500 in the KINDNESS treatment means sending back the entire amount. At first sight, it seems that the stated levels of honesty correspond at least somewhat with the actual behavior. Those who said that they would return the entire amount actually did so more often than those who said that they would keep all the money, but the difference is small and not statistically significant at conventional levels. Interestingly, almost one-third (29.6%) of the 24% who said they would keep the entire amount actually sent the entire requested amount back. Surprisingly, those who indicated in the survey that they were not sure what they would do show the highest actual return rate. It is interesting that this is the group that is most honest when it comes to actual behavior, but the absolute number of people is relatively small (20 respondents), and therefore we do not want to over-interpret the result. In our sample, 67.1% stated that they would return at least some of the money, but the actual fraction is 35.1%. Hence, we find that on average 2/3 of our participants claim to be honest, but only 1/3 actually behave in an honest way.

**Table 3: Share of those who returned the money (i.e., revealed honesty), by type of survey promise (i.e., stated honesty)**

|                             | <b>Stated honesty:<br/>Distribution</b> | <b>Fraction that returned<br/>requested amount</b> |
|-----------------------------|---|--|
| Send the entire amount back | 45.3%                                   | 37.3%  |
| Send some of money back     | 21.8%                                   | 28.6%  |
| Not sure what to do         | 8.9%                                    | 55.0%  |
| Keep the entire amount      | 24.0%                                   | 29.6%  |

Table 4 combines the data from the survey (in columns) with the data from the field experiment (in rows). For instance, 38 people who said they would keep the entire amount actually kept the entire amount; 16 of those sent some or the entire amount back. Table 4 excludes the category “not sure what to do,” because it does not exist for revealed behavior. One can see that there are deviations from the stated behavior in both directions – some people are more honest in reality compared to their stated behavior (“positive surprise”) and some are less honest (“negative surprise”) – but the second case is more frequent. If we also exclude the category “send some money back,” which is the one that is most difficult to assess in comparisons between stated and revealed behavior in our setup, we have 156 observations. Of those, 16 (10%) surprised positively and 65 (42%) surprised negatively. The remaining 75 (48%) behave as indicated in the survey.

**Table 4: Transition matrix from stated honesty to revealed honesty (excluding the survey category “Not sure what to do”)**

| <i>Stated</i>                      | <i>Keep the entire amount</i> | <i>Send some of money back</i> | <i>Send the entire amount back</i> |
|------------------------------------|-------------------------------|--------------------------------|------------------------------------|
| <b>Revealed</b>                    |                               |                                |                                    |
| <b>Keep the entire amount</b>      | 38                            | 35                             | 65                                 |
| <b>Send some money back</b>        | 16                            | 14                             | 37                                 |
| <b>Send the entire amount back</b> |                               |                                |                                    |
| Number of observations             | 54                            | 49                             | 102                                |

However, the real test of the concept is whether the stated behavior has predictive power for the actual behavior. In Table 5, we run a set of regressions to address this issue. There are several ways to look at the data. We look at the entire sample (225 observations), without excluding those for whom we are not entirely sure whether they received the money (13

observations)<sup>11</sup>. The sensitivity analysis for the smaller sample with 212 observations is provided in the appendix. We use a binary variable as the dependent variable, because all other options seem inferior for several reasons. First, the returned amount is a strangely distributed variable, which makes constructing a continuous dependent variable difficult. Second, the vast majority of participants returned either the entire requested amount or nothing. The few exceptions who returned more or less than the requested amount do not change the picture. Third, a probit regression is easier to interpret than other models such as hurdle models. To show that the results are robust – in particular, with respect to the interaction of dummy variables – we provide OLS estimates in the appendix.

Table 5 displays results that partly indicate a connection between stated and revealed behavior. The important aspect to notice is that the reference group comprises those respondents who stated in the survey that they did not know what they would do upon receiving the transfer. People who answered that they would “send some money back” or “keep the entire amount” are significantly less likely to actually return money compared to the group “I don’t know”. However, there is clearly no significant difference in the likelihood of returning money between the three main categories “send the entire amount back”, “send some money back”, and “keep the entire amount.” Put differently, the statements “keep the entire amount” and “send the entire amount back” cannot be distinguished statistically in terms of their predictive power for revealed behavior. These results are robust to the inclusion of treatment dummies and to our controls for further background variables, such as socio-economic indicators, as well as risk and ambiguity attitudes. This means that the category “Not sure what to do” stands out and that other categories are not very informative in predicting revealed behavior.

---

<sup>11</sup> This group includes those for whom, although the system sent the money to their account, for some reason we did not receive the delivery notification status of our treatment message. Because we closed our phone numbers just a few days after the experiment, we could not observe the actual delivery day of the message.

**Table 5. Regression results from probit models (marginal effects) – revealed honesty**

| <b>Dependent variable</b>              |                |                 |                 |                   |                |                   |                  |
|--|----------------|-----------------|-----------------|-------------------|----------------|-------------------|------------------|
| <b>Dummy: Money Returned</b>           | [1]            | [2]             | [3]             | [4]               | [5]            | [6]               | [7]              |
| STATED:<br>Send the entire amount back | 0.03<br>(0.06) | -               | -               | -0.17<br>(0.11)   | -              | -0.16<br>(0.11)   | -                |
| STATED:<br>Send some of money back     | -              | -0.08<br>(0.07) | -               | -0.23**<br>(0.10) | -              | -0.22**<br>(0.10) | -                |
| STATED:<br>Keep the entire amount      | -              | -               | -0.07<br>(0.07) | -0.22**<br>(0.10) | -              | -0.20*<br>(0.10)  | -                |
| STATED AMOUNT index                    | -              | -               | -               | -                 | 0.03<br>(0.08) | -                 | 0.02<br>(0.08)   |
| Treatment KINDNESS                     | -              | -               | -               | -                 | -              | 0.18**<br>(0.08)  | 0.19**<br>(0.08) |
| Treatment GUILT                        | -              | -               | -               | -                 | -              | 0.15*<br>(0.08)   | 0.14*<br>(0.08)  |
| Number of observations                 | 225            | 225             | 225             | 225               | 225            | 225               | 225              |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

We also tried to create an index for stated honesty by arbitrarily assigning the value 0 to “keep the entire amount”, 0.5 to “send some money back”, 1 to “send the entire amount back”, and 0.25 to “not sure what to do”. The index is not significant in any specification, and its coefficient remains far from conventional statistically significant levels, even if one excludes the 20 observations for the category “not sure what to do”, because their assignment to a value of 0.25 seems most arbitrary.

In general, results do not change when we interact the treatment dummies with the answer categories from the survey, although in some cases the significance of coefficients for main effects vanishes. Overall, it seems that the stated behavior has little predictive power for what people actually do in our experiment when they face the same situation in real life. To what extent is this conclusion in line with the finding in Table 4, which shows that 48% behave as stated? On average, there is some consistency, but this consistency does not necessarily lead to strong predictive power. One issue to bear in mind is the handling of the category “send some money back” (and to a lesser extent the category “not sure what to do”). Depending on how these categories are handled, the consistency level goes either up or down. The above-mentioned 48% is the upper boundary, and any other definition or assignment would reduce average consistency levels.

*Result 4: The predictive power of stated behavior for actual behavior is weak in our experiment. Indeed, there is an asymmetry: while about 70% of those who indicated that they would not send the money back stuck to their stated behavior, only about 36% of those who said that they would return the money actually did so.*

One possibility is to extend Tables 4 and 5 by regressing background and treatment variables on a dummy capturing consistency of answers regarding stated and revealed behavior. We do so in different specifications. However, it appears that our sample is not large enough to show a clear relationship between any of the background variables – be they socio-economic or preference-based – and consistency. This is not surprising. One would expect that any potential relationship is subtler and would only show up in a much larger sample.<sup>12</sup>

### 4.3 *Correlates of honesty*

Finally, we want to address the determinants of honesty. In this sub-section, we look at observable characteristics that predict whether people return money, and we disregard their stated behavior. Table 6 presents the results from probit models with a binary variable for returned money as the dependent variable. Again, the sensitivity analyses, using OLS and the restricted sample, are provided in the appendix.

Table 6 indicates that age and years of schooling are positively associated with being honest. The relationship between age and ethical behavior seems to be inversely U-shaped, indicated by the negative sign of the squared term. These results do not change when we control for income or household expenditures. Household expenditures and income are never significant in any of the regressions. We also tried to control for religious activity as a correlate, but it was always far from being statistically significant at conventional levels. We do not control for a dummy for gender as more than 90% of the respondents were male.

---

<sup>12</sup> Results are available from the authors upon request.

Finally, our prediction that uncertainty attitudes would not play a role in the decision to return money is confirmed in the data. The two coefficients for risk and ambiguity attitudes are far from being statistically significant in Model [3] in Table 6.

**Table 6. Estimation results from probit models (marginal effects) – revealed honesty: determinants.**

| <b>Dependent variable<br/>Dummy: Money Returned</b> | <b>[1]</b>         | <b>[2]</b>           | <b>[3]</b>           |
|---|--------------------|----------------------|----------------------|
| Age   | 0.000<br>(0.002)   | 0.034*<br>(0.019)    | 0.035*<br>(0.200)    |
| Age squared   | -                  | -0.0004*<br>(0.0002) | -0.0004*<br>(0.0002) |
| Years of schooling                                  | -                  | 0.021*<br>(0.012)    | 0.022*<br>(0.012)    |
| Risk aversion                                       | -                  | -                    | -0.003<br>(0.128)    |
| Ambiguity aversion                                  | -                  | -                    | -0.161<br>(0.127)    |
| First day of experiment<br>(dummy)                  | -                  | -                    | 0.053<br>(0.065)     |
| Treatment KINDNESS                                  | 0.200**<br>(0.081) | 0.205**<br>(0.082)   | 0.200**<br>(0.082)   |
| Treatment GUILT                                     | 0.143*<br>(0.082)  | 0.143*<br>(0.082)    | 0.136*<br>(0.083)    |
| <b>Number of observations</b>                       | <b>225</b>         | <b>225</b>           | <b>225</b>           |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

*Result 5: Age and years of schooling have a significant influence on the revealed level of honesty in our sample.*

The regression results in Table 6 also show the positive effect of our framing conditions. Taken together, they increased the likelihood of returning money by 60% or 16 percentage points, from 24% to 40%. After all, the inclination to behave honestly seems to be influenced quite strongly by the frame in which the decision is taken.

## 5. Discussion and Conclusion

We use a unique setting to link stated honesty to revealed honesty at the individual level in a naturally occurring situation in the field that creates a moral dilemma between being honest and accepting a monetary gain for being dishonest. Our main result indicates a strong discrepancy between stated and revealed behavior or, in other words, a weak association between words and actions.

Our results indicate that hypothetical surveys on ethical and unethical behaviors poorly reflect the actions taken in the same situations in the field. As a consequence, it seems important to use experiments, and in particular field experiments, to study the determinants of unethical behavior and to assess the effects of potential interventions to support and maintain ethical behavior. Our experiment is not the only one showing that revealed ethical and unethical behavior is quite malleable in response to the circumstances and the incentives, but it is, to our knowledge, the first to almost fully control for the comparison between stated and revealed behavior. Interestingly, we observe deviations between stated and revealed behavior in both directions, toward being more honest than self-stated and toward being less honest than stated. This indicates that there is a certain expected noise in behavior that might also be attributed to the time that passed between the survey and the experiment, but importantly there is also a clear bias. The case of stating honesty and actually behaving dishonestly is more than four times as frequent in relative terms than the opposite case. Hence, simple random variation in behavior over time cannot explain the data pattern that we observe. Probably, a quote by Groucho Marx summarizes our results neatly: “There's one way to find out if a man is honest – ask him. If he says ‘yes’ you know he is a crook.”

We also rule out explanations based on the fear of formal and informal punishment that might be differently salient in the abstract survey situation and the revealed decision. First, we argue that our institutional setup is unlikely to give rise to such fears. Second, if participants nonetheless perceived the situation as risky, participants should become more honest in the actual decision than in the hypothetical survey response. Third, controlling for risk and ambiguity attitudes as potential determinants of behavior when receiving money, which should matter if participants perceived the situation as risky or uncertain, does not change our results. The measures for uncertainty are far from being statistically significant at the conventional levels in any of our regressions.



We note that the group of participants who responded “Not sure what to do” in the survey was actually by far the most honest group when it came to actual behavior. However, they comprised slightly less than 10% of our sample. If this particular finding is robust, which has to be established in future work, one can consider well-crafted questions in surveys that might be helpful in increasing the predictability of survey answers. Surveys will always be important tools in social sciences, but one has to be careful not to over-interpret their predictive power, in particular when it comes to moral behavior.

With regard to the mechanisms implemented in our treatments, guilt works as expected, and giving a gift to encourage people to return the money does not crowd out the intrinsic inclination to return the transfer. Both have a significantly positive effect on the propensity to return the transfer, and the size of these effects is considerable. Our results are similar to Pruckner and Sausgruber (2013) in the sense that about two thirds of the sample behave dishonestly, but more importantly in terms of a significantly positive effect on honesty from changing words in a message. It seems like honest behavior can be significantly improved by minor measures such as changes in words, or as in the study by Shu et al (2012) by letting subjects sign a form with a proof of honest intent at the beginning of the form rather than at the end.

Finally, we do not want to over-interpret the levels of dishonesty in our results. Because being honest required an active choice, whereas being dishonest did not require any action, our dishonesty levels are comparatively high. There might be similar situations in other domains (e.g., when one is requested to inform authorities that one does not fulfill the criteria to receive specific social benefits anymore), but the *levels* observed in our study clearly do not carry over to all ethical decisions. Our focus was on the comparison between stated and revealed behavior and on the treatment comparisons.

## References

- Abeler, J., Becker, A., Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics* 113, 96-104.
- Alpizar, F., Martinsson, P. (2012). Paying the price of sweetening your donation – Evidence from a natural field experiment. *Economics Letters* 114, 182-185.
- Ariely, D. (2012). *The Honest Truth About Dishonesty: How We Lie to Everyone – Especially Ourselves*. Harper, New York.
- Azar, O.H., Yosef, S., Bar-Eli, M. (2013). Do customers return excessive change in a restaurant? A field experiment on dishonesty. *Journal of Economic Behavior and Organization* 93, 219-226.
- Balafoutas, L., Beck, A., Kerschbamer, R., Sutter, M. (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Review of Economic Studies* 80, 876-891.
- Bank of Tanzania (2013). *National Payment System Directorate Statistics*, <http://www.bot-tz.org/PaymentSystem/statistics.asp>.
- Becker, G.S. (1968). Crime and punishment: an economic approach. *Journal of Political Economy* 76, 169-217.
- Bertrand, M., Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *AEA Papers and Proceedings* 91, 67-72.
- Buccioli, A., Landini, F., Piovesan, M. (2013). Unethical behavior in the field: Demographic characteristics and beliefs of cheaters. *Journal of Economic Behavior and Organization* 93, 248-257.
- Cunningham, M.-R., Steinberg, J., Grev, R. (1980). Wanting to and having to help: Separate motivations for positive mood and guilt-induced helping. *Journal of Personality and Social Psychology* 38, 181-192.
- Dai, Zhixin, Galeotti, F., Villeval, M. (2016). Cheating in the lab predicts cheating in the field! An experiment in public transportation. Working Paper, GATE Lyon.
- Di Castri, S., Gidvani, L. (2014). *Enabling mobile money policies in Tanzania: A test and learn approach to enabling market-led digital financial services. Mobile Money for unbanked*. Research report submitted to Bill and Melinda Gates Foundation; available at <http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/03/Tanzania-Enabling-Mobile-Money-Policies.pdf>
- Diamond, P.A., Hausman, J.A. (1994). Contingent valuation: Is some number better than no number? *The Journal of Economic Perspectives* 8, 45-64.

- Dufwenberg, M., Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior* 30, 163-182.
- Falk, A. (2007). Gift-exchange in the field. *Econometrica* 75, 1501-1511.
- Falk, A., Heckman, J.J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535-538.
- Fehr, E., Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14, 159-181.
- Fischbacher, U., Föllmi-Heusi, F. (2013). Lies in disguise – an experimental study on cheating. *Journal of the European Economic Association* 11, 525-547.
- Franzen, A., Pointner S. (2013). The external validity of giving in the dictator game. *Experimental Economics* 16, 155-169.
- Gino, F., Ayal, S., Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization* 93, 285-292.
- Gino, F., Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science* 20, 1153-1160.
- Gneezy, U. (2005). Deception: the role of consequences. *American Economic Review* 9, 384-394.
- Haan, M., & Kooreman, P. (2002). Free riding and the provision of candy bars. *Journal of Public Economics* 83, 277-291.
- Hanemann, W.M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives* 8, 19-43.
- Hanna, R., Wang, S.-Y. (2013). Dishonesty and selection into public service. NBER Working Paper No. 19649.
- Houser, D., Vetter, S., Winter, J. (2012). Fairness and cheating. *European Economic Review* 56, 1645-1655.
- Hausman, J. (2012). Contingent valuation: from dubious to hopeless. *The Journal of Economic Perspectives* 26, 43-56.
- InterMedia (2013). *Mobile Money in Tanzania: Use, Barriers, and Opportunities*. Washington, D.C., p. 10. Available at [http://www.intermedia.org/wp-content/uploads/FITS\\_Tanzania\\_FullReport\\_final.pdf](http://www.intermedia.org/wp-content/uploads/FITS_Tanzania_FullReport_final.pdf)
- Kajackaite, A., Gneezy, U. (2016). Lying costs and incentives. Working Paper.
- Kling, C.L., Phaneuf, D.J., Zhao, J. (2012). From Exxon to BP: Has some number become better than no number? *The Journal of Economic Perspectives* 26, 3-26.

- Kröll, M., Rustagi, Devesh (2015). Got Milk? Motivations for honesty and cheating in informal markets: Evidence from India. SAFE Working Paper No. 127.
- Kube, S., Marchechal, M., Puppe, C. (2012). The currency of reciprocity: Gift-exchange at the workplace. *American Economic Review* 102, 1644-1662.
- Levitt, S. D. (2006). White-collar crime writ small: A case study of bagels, donuts, and the honor system. *American Economic Review* 96, 290-294.
- Levitt, S.D., List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives* 21, 153-174.
- List, John A. (2011). The market for charitable giving. *Journal of Economic Perspectives* 25, 157-180.
- Mellström, C., Johannesson, M. (2010). Crowding out in blood donation: Was Titmus right? *Journal of the European Economic Association* 6, 845-863.
- Potters, J., Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review* 87, 26-33.
- Pruckner, G., Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchase. *Journal of the European Economic Association* 11, 661-679.
- Rebega, O.L, Apostola, L., Benga, O. Miclea, M. (2013). Inducing guilt: A literature review. *Procedia - Social and Behavioral Sciences* 78, 536-540.
- Rosenbaum, S.M., Bilinger, S., Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth telling. *Journal of Economic Psychology* 45, 181-196.
- Shu, L.L., Mazar, N., Gino, F., Ariely, D., Bazerman, M.H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences* 109, 15197-15200.
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal*, 119, 47-60.
- Wiltermuth, S.S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes* 115, 157-168.

# Appendix

**TableA.1: Means (standard deviations) of main pre-treatment variables**

| Variable  | Full sample        | CONTROL             | KINDNESS           | GUILT              | CONTROL vs. KINDNESS <sup>a</sup> | CONTROL vs. GUILT <sup>a</sup> |
|---|--------------------|---------------------|--------------------|--------------------|-----------------------------------|--------------------------------|
| Age   | 43.41<br>(11.427)  | 43.79<br>(11.038)   | 42.25<br>(11.859)  | 44.20<br>(11.430)  | 0.39                              | 0.70                           |
| Years of schooling                                | 8.49<br>(2.727)    | 8.37<br>(2.735)     | 8.53<br>(2.952)    | 8.56<br>(2.511)    | 0.99                              | 0.72                           |
| Male dummy  | 0.93<br>(0.258)    | 0.96<br>(0.197)     | 0.93<br>(0.251)    | 0.89<br>(0.311)    | 0.47                              | 0.12                           |
| Household (HH) size                               | 4.62<br>(1.819)    | 4.75<br>(1.846)     | 4.53<br>(1.833)    | 4.59<br>(1.794)    | 0.48                              | 0.57                           |
| HH daily spending<br>(in TZS 1,000)               | 4.51<br>(2.098)    | 4.29<br>(1.986)     | 4.20<br>(2.119)    | 5.04<br>(2.111)    | 0.60                              | 0.01**                         |
| HH annual off-farm<br>income (in TZS 1,000)       | 509.68<br>(1070.9) | 449.57<br>(942.792) | 402.80<br>(691.56) | 676.66<br>(1435.7) | 0.56                              | 0.28                           |
| Experience: money sent<br>erroneously (dummy)     | 0.13<br>(0.336)    | 0.12<br>(0.327)     | 0.13<br>(0.342)    | 0.13<br>(0.342)    | 0.81                              | 0.81                           |
| Experience: money rec'd<br>erroneously (dummy)    | 0.10<br>(0.304)    | 0.09<br>(0.293)     | 0.11<br>(0.311)    | 0.11<br>(0.311)    | 0.79                              | 0.79                           |
| Alcoholic beverage daily<br>(dummy)               | 0.15<br>(0.359)    | 0.15<br>(0.356)     | 0.15<br>(0.356)    | 0.16<br>(0.369)    | 1.00                              | 0.82                           |
| Religious (dummy)                                 | 0.17<br>(0.375)    | 0.12<br>(0.327)     | 0.19<br>(0.392)    | 0.20<br>(0.403)    | 0.26                              | 0.18                           |
| Risk attitude [0,1];<br>0.5=risk neutral          | 0.56<br>(0.270)    | 0.52<br>(0.283)     | 0.57<br>(0.255)    | 0.58<br>(0.273)    | 0.30                              | 0.19                           |
| Ambiguity attitude [-1,1];<br>0=ambiguity neutral | 0.00<br>(0.282)    | 0.03<br>(0.339)     | -0.01<br>(0.253)   | -0.01<br>(0.249)   | 0.13                              | 0.76                           |
| Observations                                      | 225                | 75                  | 75                 | 75                 | 75                                | 75                             |

<sup>a</sup> P-values for two-sided Mann-Whitney-U-test.

\*\* significant on the 1%-level.

**Table 5a. Regression results from probit models (marginal effects) – revealed honesty  
(212 observations)**

| <b>Dependent variable<br/>Dummy: Money Returned</b> | <b>[1]</b>     | <b>[2]</b>       | <b>[3]</b>        | <b>[4]</b>         | <b>[5]</b>     | <b>[6]</b>       | <b>[7]</b>       |
|---|----------------|------------------|-------------------|--------------------|----------------|------------------|------------------|
| STATED:<br>Send the entire amount back              | 0.02<br>(0.07) | -                | -                 | -0.20*<br>(0.11)   | -              | -0.20*<br>(0.11) | -                |
| STATED:<br>Send some of money back                  | -              | -0.09<br>(0.088) | -                 | - 0.26**<br>(0.10) | -              | 0.26**<br>(0.10) | -                |
| STATED:<br>Keep the entire amount                   | -              | -                | - 0.045<br>(0.08) | -0.23**<br>(0.10)  | -              | -0.21*<br>(0.11) | -                |
| STATED AMOUNT index                                 | -              | -                | -                 | -                  | 0.01<br>(0.08) | -                | -0.01<br>(0.08)  |
| Treatment KINDNESS                                  | -              | -                | -                 | -                  | -              | 0.19**<br>(0.09) | 0.20**<br>(0.08) |
| Treatment GUILT                                     | -              | -                | -                 | -                  | -              | 0.18**<br>(0.09) | 0.17**<br>(0.08) |
| Number of observations                              | 212            | 212              | 212               | 212                | 212            | 212              | 212              |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

**Table 5b. Regression results from OLS models– revealed honesty  
(full sample)**

| <b>Dependent variable<br/>Dummy: Money Returned</b> | <b>[1]</b>     | <b>[2]</b>      | <b>[3]</b>      | <b>[4]</b>        | <b>[5]</b>     | <b>[6]</b>       | <b>[7]</b>       |
|---|----------------|-----------------|-----------------|-------------------|----------------|------------------|------------------|
| STATED:<br>Send the entire amount back              | 0.03<br>(0.06) | -               | -               | -0.19<br>(0.12)   | -              | -0.18<br>(0.12)  | -                |
| STATED:<br>Send some of money back                  | -              | -0.08<br>(0.08) | -               | -0.26**<br>(0.13) | -              | -0.25*<br>(0.13) | -                |
| STATED:<br>Keep the entire amount                   | -              | -               | -0.07<br>(0.07) | -0.25**<br>(0.12) | -              | -0.22*<br>(0.12) | -                |
| STATED AMOUNT index                                 | -              | -               | -               | -                 | 0.03<br>(0.08) | -                | 0.01<br>(0.08)   |
| Treatment KINDNESS                                  | -              | -               | -               | -                 | -              | 0.17**<br>(0.08) | 0.19**<br>(0.08) |
| Treatment GUILT                                     | -              | -               | -               | -                 | -              | 0.14*<br>(0.08)  | 0.13*<br>(0.08)  |
| Number of observations                              | 225            | 225             | 225             | 225               | 225            | 225              | 225              |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

**Table 5c. Regression results from OLS model – revealed honesty  
(212 observations)**

| <b>Dependent variable<br/>Dummy: Money Returned</b> | <b>[1]</b>     | <b>[2]</b>      | <b>[3]</b>      | <b>[4]</b>        | <b>[5]</b>     | <b>[6]</b>        | <b>[7]</b>       |
|---|----------------|-----------------|-----------------|-------------------|----------------|-------------------|------------------|
| STATED:<br>Send the entire amount back              | 0.02<br>(0.07) | -               | -               | -0.22*<br>(0.12)  | -              | -0.21*<br>(0.12)  | -                |
| STATED:<br>Send some of money back                  | -              | -0.09<br>(0.09) | -               | -0.30**<br>(0.13) | -              | -0.29**<br>(0.13) | -                |
| STATED:<br>Keep the entire amount                   | -              | -               | -0.05<br>(0.08) | -0.27**<br>(0.13) | -              | -0.24*<br>(0.13)  | -                |
| STATED AMOUNT index                                 | -              | -               | -               | -                 | 0.01<br>(0.08) | -                 | -0.01<br>(0.08)  |
| Treatment KINDNESS                                  | -              | -               | -               | -                 | -              | 0.17**<br>(0.08)  | 0.19**<br>(0.08) |
| Treatment GUILT                                     | -              | -               | -               | -                 | -              | 0.16**<br>(0.08)  | 0.16**<br>(0.08) |
| Number of observations                              | 212            | 212             | 212             | 212               | 212            | 212               | 212              |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

**Table 6a. Regression results from probit models (marginal effects) –  
revealed honesty: correlates (212 observations)**

| <b>Dependent variable<br/>Dummy: Money Returned</b> | <b>[1]</b>         | <b>[2]</b>           | <b>[3]</b>           |
|---|--------------------|----------------------|----------------------|
| Age   | 0.001<br>(0.003)   | 0.035*<br>(0.021)    | 0.036*<br>(0.207)    |
| Age squared   | -                  | -0.0004*<br>(0.0002) | -0.0004*<br>(0.0002) |
| Years of schooling                                  | -                  | 0.021<br>(0.013)     | 0.022*<br>(0.013)    |
| Risk aversion                                       | -                  | -                    | -0.012<br>(0.132)    |
| Ambiguity aversion                                  | -                  | -                    | -0.210<br>(0.134)    |
| Treatment KINDNESS                                  | 0.203**<br>(0.084) | 0.204**<br>(0.085)   | 0.203**<br>(0.086)   |
| Treatment GUILT                                     | 0.169**<br>(0.085) | 0.163*<br>(0.085)    | 0.162*<br>(0.086)    |
| Number of observations                              | 212                | 212                  | 212                  |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

**Table 6b. Regression results from OLS models –  
revealed honesty: (full sample)**

| <b>Dependent variable</b><br><b>Dummy: Money Returned</b> | <b>[1]</b>         | <b>[2]</b>           | <b>[3]</b>           |
|---|--------------------|----------------------|----------------------|
| Age   | 0.001<br>(0.003)   | 0.031*<br>(0.018)    | 0.032*<br>(0.018)    |
| Age squared   | -                  | -0.0003*<br>(0.0002) | -0.0003*<br>(0.0002) |
| Years of schooling  | -                  | 0.019*<br>(0.012)    | 0.020*<br>(0.012)    |
| Risk aversion   | -                  | -                    | -0.010<br>(0.126)    |
| Ambiguity aversion  | -                  | -                    | -0.147<br>(0.120)    |
| Treatment KINDNESS  | 0.188**<br>(0.077) | 0.194**<br>(0.077)   | 0.190**<br>(0.077)   |
| Treatment GUILT   | 0.133*<br>(0.077)  | 0.132*<br>(0.077)    | 0.126<br>(0.077)     |
| Number of observations                                    | 225                | 225                  | 225                  |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.

**Table 6c. Regression results from OLS models –  
revealed honesty: correlates (212 observations)**

| <b>Dependent variable</b><br><b>Dummy: Money Returned</b> | <b>[1]</b>         | <b>[2]</b>          | <b>[3]</b>           |
|---|--------------------|---------------------|----------------------|
| Age   | 0.001<br>(0.003)   | 0.032*<br>(0.019)   | 0.032*<br>(0.019)    |
| Age squared   | -                  | -0.0003<br>(0.0002) | -0.0003*<br>(0.0002) |
| Years of schooling  | -                  | 0.019<br>(0.012)    | 0.020*<br>(0.012)    |
| Risk aversion   | -                  | -                   | -0.013<br>(0.129)    |
| Ambiguity aversion  | -                  | -                   | -0.184<br>(0.123)    |
| Treatment KINDNESS  | 0.191**<br>(0.080) | 0.191**<br>(0.080)  | 0.186**<br>(0.080)   |
| Treatment GUILT   | 0.157*<br>(0.080)  | 0.150*<br>(0.080)   | 0.144*<br>(0.080)    |
| Number of observations                                    | 212                | 212                 | 212                  |

Note: \*\*\* denotes significance at the 1% level, \*\* at the 5% level and \* at the 10% level. Standard errors in parentheses.