

The external validity of giving in the dictator game

A field experiment using the misdirected letter technique

Axel Franzen · Sonja Pointner

Received: 31 January 2012 / Accepted: 18 August 2012 / Published online: 1 September 2012
© Economic Science Association 2012

Abstract We investigate the external validity of giving in the dictator game by using the misdirected letter technique in a within-subject design. First, subjects participated in standard dictator games (double blind) conducted in labs in two different studies. Second, after four to five weeks (study 1) or two years (study 2), we delivered prepared letters to the same subjects. The envelopes and the contents of the letters were designed to create the impression that they were misdirected by the mail delivery service. The letters contained 10 Euros (20 Swiss Francs in study 2) corresponding to the endowment of the in-lab experiments. We observe in both studies that subjects who showed other-regarding behavior in the lab returned the misdirected letters more often than subjects giving nothing, suggesting that in-lab behavior is related to behavior in the field.

Keywords Other-regarding preferences · Fairness · Dictator game · Behavioral game theory · External validity · Misdirected letter technique · Field experiment

JEL Classification C91 · C93 · D03 · D63 · D64

1 Introduction

Natural field experiments are still rare in social science, but most scholars consent that they can offer important complementary insight to laboratory studies (e.g. Andreoni et al. 2011; Card et al. 2011; Harrison and List 2004; Levitt and List 2007, 2009;

Electronic supplementary material The online version of this article (doi:10.1007/s10683-012-9337-5) contains supplementary material, which is available to authorized users.

A. Franzen (✉) · S. Pointner
Faculty of Economics and Social Sciences, Institute of Sociology, University of Bern,
Lerchenweg 36, 3012 Bern, Switzerland
e-mail: axel.franzen@soz.unibe.ch

List 2011). Experiments conducted in the lab have many advantages, such as the ability to control the stimulus and conduct randomization of subjects between the control and treatment group. Therefore, laboratory studies are considered to be the gold standard of scientific inquiry. However, they also have the downside of often being rather abstract and remote from the real life experiences of subjects. Therefore, many laboratory studies investigating decision behavior leave us in doubt as to what extent the results can be generalized to real world behavior.

Levitt and List (2007) point out a number of differences between laboratory settings and naturally occurring situations. First, laboratory studies are mostly conducted with student volunteers, who might be a very special group of subjects. Lab participants self-select into the group of university students, and furthermore self-select into the group of volunteers. They are younger and better educated than the general population and might also differ on other important unobserved characteristics (e.g. risk seeking, cooperativeness). This raises the question of whether findings based on laboratory studies can be generalized to other individuals.

Second, for a number of reasons, the decision situations presented in the lab are usually abstract and unparalleled in real life situations. Context free decision situations have the advantage that uncontrolled differences in subjects' experiences or interpretation of the experimental setup do not distort experimental results and therefore enable better control of the internal stimulus. However, this raises the question of whether results from artificial decision tasks can be generalized to behavior outside the lab. Third, subjects might be more generous in experiments because of "windfall gains", whereas outside the lab they have to deal with earned money (Cherry et al. 2002). Fourth, and maybe most importantly, subjects in the lab are aware that they are scrutinized, and it is often unclear how much of their behavior can be explained by their true preferences or by attempts to comply with the expectations of the researcher or to signal cooperativeness and fairness to assumed observers.

All of the above caveats refer particularly to experiments designed to measure other-regarding preferences, such as the dictator game, ultimatum game, or contribution behavior in public good games. For instance, more than a hundred experiments with the dictator game show that individuals share substantial amounts (on average 28 %) of their endowments with anonymous recipients (see Engel 2011 for a meta analysis). If these results have external validity, they would present an anomaly to the standard economic assumption of monetary payoff maximization. However, there are also a number of experimental studies casting doubt on the interpretation that subjects share resources because of other-regarding preferences. Thus, Bardsley (2008) and List (2007) find that individuals share much less or even withdraw money if dictators have the option to do so. Dana et al. (2007) as well as Andreoni and Bernheim (2009) show that giving is reduced if the outcome is co-determined by chance. Franzen and Pointner (2012) demonstrate that giving almost disappears if decisions are concealed from recipients and the experimenter by using the randomized response technique. However, also these lab experiments can be critiqued on the grounds of lacking external validity and displaying subtle experimenter demand effects.

In this paper we consider whether the standard dictator game bears any relevance to the behavior of individuals in a natural field setting. To test the external validity displayed in a dictator game, we conducted two studies where in each case the same

subjects are observed some time after the experiment. We use the misdirected letter technique, which allows us to observe subjects in a situation that is similar to the decision setup in the lab without them being aware of it. Our studies are similar to the studies by Benz and Meier (2008), Baran et al. (2009), and Englmaier and Gebhardt (2010) who also used within-subject designs. Thus, our study does not address the selectivity argument of student populations, but it eliminates the experimenter demand effect, because subjects are tested in a real life context.

The remainder of the paper is organized into three sections. The next paragraph shortly reviews the literature that is most relevant to our study. Particularly, we discuss similar field studies in which the same subjects are observed in and outside the lab. The third section describes the design of our own lab and field studies. We use the misdirected letter technique in our field experiment, and because the method is not well known, we describe it in more detail. The section also contains the description of our results. Finally, the last section summarizes and discusses the results.

2 Evidence from natural field experiments

So far, there are only a handful of studies investigating other-regarding preferences by comparing individuals' behavior in and outside the lab via a natural field experiment.¹ The studies that are closely related to our experiment are Benz and Meier (2008), Baran et al. (2009) and Englmaier and Gebhardt (2010). In Benz and Meier's (2008) study, students at the University of Zurich are asked every semester by the university's enrollment office whether they want to donate money (about 12 Swiss Francs or about \$12) to the university's "social fund", which supports students in need. Benz and Meier (2008) were able to obtain the data and conduct two in-class experiments giving participants the same amount and asking how much they would be willing to donate to the same student fund (study 1) or to another fund not related to the university (study 2). They find correlations from 0.25 to 0.4 between subjects' in-class behavior and their response to the university's donation request depending on whether out-of-class behavior is compared before or after the in-class experiment. However, both in-class and out-of-class situations are very similar and the chance that students were able to guess the purpose of the experiment seems high. Thus, it is unclear to what extent the results are due to students' attempt to be consistent. The fact that the authors use an in-class experiment also limits the anonymity of the situation since there is a high chance that students discuss their behavior after class. The low anonymity is also reflected in the fact that a very high proportion of students (80 % in study 1 and 65 % in study 2) donated some of their endowments in the laboratory study.

¹ Generally, studies dealing with field experiments are still rare. Card et al. (2011) counted studies reporting the results of field experiments that were published in one of the top five economic journals (AER, ECON, JPE, QJE or RES) during the last 35 years (1975 to 2010). They found 84 studies overall, most of them (particular until 2005) in the field of labor economics. Only eight (starting with a study by Frey and Meier 2004) are dealing with other-regarding preferences or cooperation.

Baran et al. (2009) conducted trust games with MBA students at the University of Chicago Business School to observe subjects' response to the university's alumni donation campaign at the time of graduation. They report a correlation of 0.29 between the proportion subjects returned as trustees in the lab experiments and the donations made to the business school. Thus, although the design of the study differs from Benz and Meier (2008) with respect to the experiment (trust game) and field setting (much larger time lag), the results are very similar and confirm external validity of reciprocal behavior displayed in the experiment. However, both studies are similar with respect to the high level of reciprocal behavior observed in the lab (87.9 % return something in the trust game) and the high level of contribution to the alumni campaign (94.8 % donate at least a dollar). Thus, both studies use field environments with high return rates, which suggests low anonymity in the lab experiments and in the field.

Englmaier and Gebhardt (2010) conducted field experiments with students at the University of Munich. Their subjects had to register books in the university's library in groups of four without knowing that they were being observed. Two weeks later, some of the same subjects participated in a standard public good experiment. The authors find that subjects who showed less free riding in the library task also displayed less free riding behavior in the lab. The reported correlation between the field setting and the contribution behavior in the lab is 0.4.

There are also some studies that test external validity using non-student subjects. Karlan (2005) conducted trust games and public good games with indigenous inhabitants of Peru to observe their repayment behavior of micro loans. He finds that the behavior in the trust game is correlated with behavior in the field but no consistent results were obtained for the public good game. Fehr and Leibbrandt (2008) conducted public good games with fishermen in Brazil. They find that those who contributed more in the lab experiment also exploited common fishing grounds less by using larger fishnet meshes. Barr and Zeitlin (2010) analyzed teachers in Uganda. The time they spent working (lecturing in schools) is positively but weakly correlated with giving behavior in a dictator game. However, there are also some studies reporting no or only partly consistent behavior between the lab and the field. Hence, Carpenter and Meyers (2010) show that volunteer firefighters give more in a dictator game compared with their non-volunteer colleagues. But giving does not correlate with the amount of fire-fighting operations. Gurven and Winking (2008) find no relation between the in-lab experimental behavior (dictator and ultimatum games) of Bolivian indigenous inhabitants and the observed every day sharing of food.

Our study is also closely related to a recent experiment conducted by Stoop (2012). Stoop (2012) tests for experimental setting and scrutiny effects in a between-subjects design administering misdirected letters to subjects in Tilburg (Netherlands). A letter containing 10 Euros and addressed to a volunteer is given to students in a lab, to ordinary citizens in a lab and at home, and is dropped into the mail boxes of randomly chosen households. Subjects have the choice to either keep the letter or to forward it to the addressed person. In all four groups the letters are returned by roughly the same rate of 50 %. Hence, the results suggest that subjects behave similarly in the lab as in the field and that student volunteers do not differ from ordinary citizens.

Our study differs from the preceding ones in that we conduct a standard dictator game using strict double blind conditions. Moreover, our field studies present a

high level of anonymity by using the misdirected letter technique, which allows us to keep the stakes involved in the lab experiment and in the field constant. Hence, the incentive structure is literally identical in the lab as in the field. In contrast to Stoop (2012), we use a within-subject design. Therefore, our study answers the question of how giving in the dictator game is related to the response given to a misdirected letter. The design of our lab experiment and the field study are described in detail in the next section.

3 The design of the laboratory and field experiment

Our first study was conducted in May 2009 at the University of Cologne. We recruited students on campus who were invited to the lab where they played a dictator game. The in-lab experiment was conducted similarly to the double blind procedure used by Hoffman et al. (1996). Thus, subjects' decisions remained unknown to the experimenter as well as to the recipients. Participants received an endowment of 10 Euros and were told that they could divide the 10 Euros between themselves and the other person. It was made explicit that every allocation is possible, including keeping or giving everything. The laboratory was equipped with experimental booths with written instructions for each subject, a large envelope containing ten one Euro vouchers, and two small envelopes, one red and one blue. Subjects were instructed to put the Euro vouchers they wanted to keep into the red envelope and the ones they wanted to give away into the blue envelope. They had to place both envelopes in the larger envelope that was afterwards dropped into a box. A student assistant took the box to another room. For payout, the amount of vouchers in the red envelope was changed into real money. The money was enclosed in small envelopes and placed on a desk. Subjects passed this room on their way to the exit and could pick up their envelopes without seeing anyone. It was explained to the subjects that the procedure of placing the vouchers into envelopes guaranteed the anonymity of their decision.

To test the external validity of the laboratory experiment we use the “misdirected-letter-technique” (MLT). MLT is a variation of the well-known “lost-letter-technique” (LLT), which was introduced by Merritt and Fowler (1948). LLT has been successfully utilized in empirical research for non-reactive measurements of attitudes (see Milgram et al. 1965; Milgram 1969; Sechrest and Belew 1983; Bridges et al. 2001; Wilson et al. 2009). However, LLT has some shortcomings. One disadvantage is the fact that the researcher has little control over who finds the letter. Particularly, it is usually impossible to control whether single persons or groups find the letter or to control for other circumstances (e.g. weather conditions), which could influence the return rates. Since the researcher has no control over the composition of the finder population, the sample is not representative and possibly biased towards some undesirable characteristics (e.g. unemployed persons, elderly people). A further problem is that curious finders sometimes open lost letters, which makes returning less likely.²

²Letters to a “Sex Research Project” have been opened more frequently than letters to an “Education Research Project” (Sechrest and Belew 1983).

Furthermore, LLT is a well-known technique and often the subject of lectures in social science classes, which increases the likelihood that student subjects could be familiar with the technique.

MLT addresses these problems. Most importantly, it is unknown to students because it has been rarely applied so far. We are aware of only three studies utilizing MLT (Howitt et al. 1977; Howitt and McCabe 1978; Kremer et al. 1986). In contrast to LLT, the letters in MLT are not “lost” but intentionally “misdirected” to the address of a specific household. This allows for greater control of the subjects’ characteristics. In a study investigating attitudes towards Asian immigrants in the UK, Howitt et al. (1977) misdirected letters to households with English family names and to households suggesting an Asian background. Both types of households received letters with real addresses but fake names such as “Mr. Edward” and “Mr. Singh”. Households with the Asian background returned both letters equally. However, the English households returned the letters addressed to the “Mr. Edwards” at a rate of 88 %, but the letters addressed to “Mr. Singhs” at a rate of only 55 %, suggesting some reservations towards Asian immigrants. In a similar study, Howitt and McCabe (1978) investigated attitudes of British inhabitants towards the Irish. Again, the return rates of letters to “Mr. O’Reilly” (an Irish name) and “Mr. Bell” (an English name) differed in the assumed direction by 46 percentage points. Kremer et al. (1986) tested the attitudes of Catholics towards Protestants and vice versa in Northern Ireland. Letters were misdirected to Catholic and Protestant households with the names of “Patrick Connolly” (a Catholic name) and “William Scott” (a Protestant name). Catholic households returned only 15 % of letters to the Protestant name while Protestant households reacted equally towards “Connolly” and “Scott”.

We use MLT to test the external validity of giving in the dictator game. Directly following the lab experiment we asked our subjects in an online-questionnaire to leave their address if they would consider participating in further experiments. The questionnaire contained a code which was not visible to subjects but allowed us to link allocations made in the dictator game to addresses (if given). Overall, 179 of 276 subjects that participated in the experiment left the address information. In the questionnaire, participants were also asked about the size of the household they live in. In order to make sure that only subjects who participated in the experiment find the misdirected letters, the letters were only delivered to subjects living in single person households. This provided us with 27 subjects who finally received the letter in study 1.

One difficulty when using MLT is the delivery of letters in such a way, as subjects cannot discover their involvement in an experiment. Kremer et al. (1986) folded the letter so that the addressees’ names were partly unreadable and only the rest of the (real) address was visible. However, this method only allowed the delivery of one-third of the letters by the postal service. In order to avoid the risk that the mail service discovers the wrong addresses and does not deliver the letters, we decided to distribute the envelopes ourselves. For this reason, the letters contained a slight variation of the real address of the target household, the real family name, and an identical surname initial. This way subjects had to open the letters in order to find out that they were not the intended receivers, since a different surname appeared in full

length in the letter.³ Letters were distributed during the standard delivery time of the post service. Furthermore, to make letters look real, we had them stamped with the official postal date stamp.⁴

The subjects received the letters about 4 to 5 weeks after the laboratory experiment while the semester was still going on. Therefore, our subjects should have been available at their home address. Moreover, the likelihood of an individual moving to a different address during this short time period is very low. The letter contained the slightly manipulated address and a short text. The text was designed to give subjects the impression that it was written by a grandmother to congratulate her grandchild for good school grades: “Dear Alexander, we congratulate you for your excellent grades. Grandfather and I are very proud of you. Best regards, your grandma Sophia”. A copy of an original letter in German is displayed in Fig. 1. As a present for the grandchild, the letter contained 10 Euros but this was not mentioned in the letter.

4 Results

The distribution of giving in the lab experiment is displayed in Fig. 2. On average, the 27 subjects gave 2.51 Euros to recipients. The distribution shows that 9 subjects (33 %) gave nothing and 5 persons divided their amount by half. The allocations in our experiment are comparable to other distributions with simple dictator games and standard conditions of anonymity (see Camerer 2003; Engel 2011). The distribution is also not significantly different from the other 249 subjects who did not receive a misdirected letter.⁵ Thus, there seems to be no selectivity with respect to other-regarding preferences in the subsample of the 27.

Of the 27 misdirected letters 14 were forwarded and returned to the sender. All of the returned letters were opened by the recipients and all of them still contained the 10 Euro note. Three of the returned letters also contained additional notes written by subjects stating that the letter was obviously misdirected. Thus, we have no indication that subjects discovered their involvement in a field experiment.

Next, we consider whether subjects who displayed fairness in the lab also show fairness in the field. The 14 subjects who returned the letter allocated on average 3.36 Euros to co-players, while those who did not return the letter donated 1.62 Euros. A *t*-test results in a *t*-value of $t = 1.76$ which is statistically significant for a one-tailed test ($p = 0.045$) at the 5 % significance level and on the 10 % level for a two-sided test ($p = 0.09$) (compare Table 2). Furthermore, we classify participants as regarding

³The addresses were scattered throughout the urban area of Cologne (Germany).

⁴These stamps cannot be received on request. Rather we got hold of officially stamped envelopes by asking a colleague from a neighboring city (Bonn) to mail us unsealed envelopes a day before the experiment. We were able to reuse these envelopes for the experiment. The sender address of the misdirected letter came from the same city the official stamp came from (Bonn) so that letters looked consistent and real. The fact that the sender address and the stamp come from a different city than subjects' place of residence was done purposely to avoid that subjects could carry letters in person to the sender's address and to dilute any suspicion that the University of Cologne was involved.

⁵From the 249 subjects 29 % kept everything and the average given to others is 2.94 Euro. The distributions do not differ according to a Kolmogorov-Smirnov test as well as to a standard *t*-test.

Table 1 The association between experimental behavior and field behavior

	Fair in dictator game		
	not fair	fair	
letter not returned	6 66.7%	7 38.9%	13 48%
letter returned	3 33.3%	11 61.1%	14 52%
Total	9	18	27

Note: $\chi^2=1.854$ with $p=0.173$, $\phi=0.262$

May 2010 and stem from three different groups. One group played the dictator game exactly as in study 1. In the second group, subjects first played a different version of the dictator game and afterwards (though knowingly unrelated) the standard double blind dictator game (as in study 1). In a third group, subjects were provided with the pictures of the recipients. However, comparing the distributions of giving in the three groups revealed no significant differences. Thus, we are confident in pooling the data from the three groups, which results in an overall sample of 233 subjects. As in study 1, students answered a questionnaire after the experiments and could leave an address if they wanted to participate in further experiments, which 165 of the 233 did. Of these 165 some addresses were no longer valid or subjects left Bern by the time we conducted the field experiment. Finally, we ended up with 77 addresses to which we delivered a misdirected letter in May 2012. In contrast to study 1, letters were also delivered to subjects living in multiple person households. The letter was designed as in study 1, but contained 20 Swiss Francs (as the endowment in the dictator games in study 2) as opposed to 10 Euros. The delivery of the misdirected letters was also done as in study 1. From the 77 letters that we slipped into the mailboxes of subjects, 55 were returned. Most of the letters were opened and returned by subjects in an extra envelope accompanied by a note, stating that the letter was obviously misdirected. However, two subjects returned the letters stating that they suspected being part of an experiment. Both returns did not include the 20 Swiss Francs. Rather, the subjects stated that they would forward the money if they would receive the correct address, accounting obviously for the insecurity about their guess. We deleted both cases from further analysis.

Figure 3 displays the distribution of giving in the dictator game in Bern. The allocation of the 75 subjects in Bern who received the letter does not differ from the distribution of the rest of the subjects that participated in the dictator games according to a Kolmogorov-Smirnov test.

Second, we compare the allocations from the in-lab experiment in Bern ($N = 75$) with the allocations in Cologne ($N = 27$). Although our Bern subjects were a little more generous than the subjects in Cologne (25.3 % gave nothing as compared to

Fig. 3 Allocations of subjects in the dictator game of study 2 ($N = 75, \bar{x} = 5.17$)

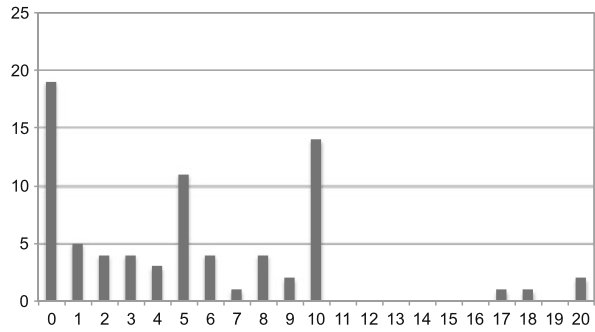


Table 2 Comparison of mean allocations (returned versus non-returned letters)

	means: not returned/ returned	difference means	<i>t</i> -statistic	<i>p</i> -value: one-sided	<i>p</i> -value: two-sided
Study 1 (Cologne, $n = 27$)	1.62 ($n = 13$)/ 3.36 ($n = 14$)	-1.74	-1.76	0.045	0.090
Study 2 (Bern, $n = 75$)	3.59 ($n = 22$)/ 5.83 ($n = 53$)	-2.24	-1.81	0.037	0.073

33.3 %), the test statistics indicate no significant differences. Hence, the in-lab results of study 1 and study 2 are not significantly different.

The 53 subjects who returned the letter gave on average 5.83 Swiss Francs to the recipient in the in-lab dictator game, while the average was only 3.59 for those who did not return it. The *t*-value equals 1.81 and is statistically significant for a one-sided test ($p = 0.037$) and on the 10 % level for a two-sided test ($p = 0.07$) (Table 2). Moreover, the phi correlation between giving something in the lab (versus giving nothing) and returning the letter (versus not returning) equals 0.1634 and is statistically not different from the correlation of study 1 (Fisher’s *z*-test: $z = 0.43, p = 0.67$). Thus, study 2 replicates the results of study 1.

Next we pool the data from study 1 and study 2. The resulting cross-tabulation is presented in Table 3. A χ^2 -test ($\chi^2 = 4.21$) indicates that subjects who gave some money to recipients in the experiment also returned the misdirected letter more often ($p = 0.04$). However, the relationship is not very strong. 72 % of the subjects who gave some money to recipients also returned the letter compared to 50 % who gave nothing, resulting in a difference of 22 % percentage points or a phi correlation of 0.2.

Since we gathered further information about subjects in the online-questionnaires in both studies, we can also control for subjects’ socio-demographic characteristics and the place where the experiments were conducted. Furthermore, we were interested in whether subjects’ level of trust, their social network size, and social desirability are related to returning the letter.⁶ One explanation of why people behave

⁶See Appendix for a table with descriptive statistics of the variables.

Table 3 The association between lab behavior and field behavior of study 1 and 2

	Fair in dictator game		
	not fair	fair	
letter not returned	14 50.0%	21 28.4 %	35 34.3%
letter returned	14 50.0%	53 71.6 %	67 65.7 %
total	28 100%	74 100%	102

Note: $\chi^2=4.213$ with $p=0.040$, $\phi=0.203$

fairly is that they like to signal fairness (Andreoni and Bernheim 2009). In this case, subjects with larger networks could have a higher motivation to signal fairness as compared to those with small networks. Similarly, the extent of trust in others could be an indicator for the importance or closeness of others and thus could increase the motive to signal fairness. In addition, the questionnaire contains a measure of subjects' social desirability, which could explain their conformity to social norms. The results of the multivariate logit analyses are presented in Table 4. Model 1 presents the bivariate estimate of giving something in the dictator game and the probability of returning the letter. The logit coefficient is positive and significant at the 5 % level, echoing the result obtained from the cross-tabulation. The odds of returning the letter increase by 2.5 if subjects gave some money to the recipient in the experiment (i.e. the probability for returning the letter increases by 0.22).

In Model 2 we introduce subjects' gender, age, income, the place where the study was conducted, and whether subjects live in a single household or not.⁷ None of these control variables have a significant influence on behavior in the field, and they do not alter the effect of giving in the lab. Finally, in Model 3 we add further measures of subjects' trust in other people, network size, and a measure of social desirability. Our trust variable is an index consisting of answers to three questions that are also used in well-known surveys (e.g. WVS) (see Appendix for question wording). Network size was measured by asking subjects about the number of weak and closely related friends. Social desirability is measured by using a short version (four items, see Appendix) of the well-known Marlowe-Crowne Scale (Crowne and Marlowe 1960). The results show that none of these control variables have a statistically significant influence on the observed field behavior, and the effect of giving in the lab remains stable.

⁷In Cologne the letter was only delivered to subjects living in single households. In the Bern study 16 of all 75 participants lived in single households.

Table 4 Logit regression, dependent variable “letter returned” (coded 1)

letter returned (dependent variable)	Model		
	(1)	(2)	(3)
fairness in DG (fair = 1)	0.926* (2.02)	0.960* (2.00)	1.017* (2.02)
sex (female = 1)		0.189 (0.40)	0.211 (0.44)
age		-0.108 (-1.57)	-0.109 (-1.59)
income (z-standardized)		0.341 (1.32)	0.330 (1.27)
household (single = 1)		-0.441 (-0.66)	-0.377 (-0.55)
Bern (study 2 = 1)		0.365 (0.52)	0.412 (0.55)
trust in people			-0.014 (-0.13)
social desirability			-0.171 (-0.35)
number of friends			0.003 (0.34)
Constant	-6.32e-16 (-0.00)	2.326 (1.28)	2.341 (1.13)
Observations	102	102	102
McFadden R ²	0.031	0.093	0.095

z-value in parentheses;

* $p < 0.05$

5 Discussion and conclusion

This paper reports the results of two studies that were conducted at the Universities of Cologne ($N = 27$) and Bern ($N = 75$). In both studies we first conducted standard double blind dictator games. In Cologne, the same subjects received a misdirected letter four to five weeks after the lab experiment. In Bern, the misdirected letter was delivered two years after the lab experiment. In order to keep the incentive structure constant, the letters contained the same amount as was at stake in the experiments, 10 Euros in Cologne and 20 Swiss Francs in Bern. Both studies show very similar results in the lab as well as in the field. In Cologne, 33 % of our subjects kept the whole endowment in the lab. Of those who gave some money to recipients in the lab, 61 % also returned the letter in the field whereas only 33.3 % returned it if nothing was allocated to the recipient. In Bern, 25 % kept everything in the dictator game. Of those who gave some money in the lab, 75 % returned the letter in the field as compared to 58 % who shared nothing in the lab. The difference in terms of percentage points (28 % and 17 %) or correlation ($\phi = 0.26$ and 0.16) between the two studies is small and statistically not significant. Hence, study 2 replicates the findings of study 1, even

though the misdirected letter was delivered two years after the experiment in study 2. If we combine the two samples, then the misdirected letter was returned by 72 % who gave something as compared to 50 % who shared nothing in the lab, resulting in a difference of 22 percentage points or a phi correlation of 0.20. Our results are closely in line with other studies using within-subject designs (Benz and Meier 2008; Baran et al. 2009; Englmaier and Gebhardt 2010). In sum, the misdirected letter technique, as used in the field, confirms the hypothesis that the decision situation of the dictator game has real life relevance.

Our results, as well as results from other field experiments, suggest that the association is only weak or moderate. However, whether the glass is half full or half empty depends on one's expectations. Although we tried to find a field situation that closely resembles the dictator game, the two setups differ.⁸ The misdirected letter was addressed to a specific person (as compared to a randomly assigned recipient in the lab) and, therefore, the money was "earned" by the recipient in the field. On one hand, this should elicit more prosocial reactions in the field than in the lab. On the other hand, forwarding a misdirected letter takes more effort than allocating something in a dictator experiment. This should lead to lower return rates. Given these differences between the lab and field experiment, a correlation of 0.2 makes the glass look half full.

Stoop (2012) finds no difference in results if the misdirected letter technique is used in or outside the lab. However, Stoop uses a between-subject design and therefore compares group averages. If we compare the group margins of our studies, then the findings are similar to Stoop's: 65.7 % returned the letter and 72.5 % gave something in the lab. However, the within-subject design reveals that only a subsample of subjects behave prosocially in the lab as well as in the field. Consistently prosocial behavior in both situations is shown by 52 % of subjects and consistently selfish behavior by 14 %. The rest of the subjects switch decisions depending on the situation.⁹

It should be emphasized that our results do not show that the consistent half of our subjects really have other-regarding preferences. The results could also be driven by their attempt to appear fair in the eyes of possible observers. We tried to implement high anonymity of the decisions inside and outside the lab. However, evidence from other studies demonstrates that standard double blind procedures render less anonymity than methods with a random mechanism such as the randomized response technique (Franzen and Pointner 2012) or decision concealment by using a chance mechanism (Dana et al. 2007; Andreoni and Bernheim 2009). Also, returning a misdirected letter is not necessarily done in isolation. Subjects might feel observed by other household members, friends or the postal service. Therefore, we are careful in interpreting the results: Patterns of giving in and outside the lab could also be driven by individuals' intention to signal fairness.

⁸Field settings will probably always differ in some aspect from in-lab experiments, which limits the expected correlations (see Camerer 2011).

⁹See also Blanco et al. (2011) who find inconsistencies of subjects' behavior in different games.

References

- Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607–1636.
- Andreoni, J., Justin, M. R., & Trachtman, H. (2011). *Avoiding the ask: a field experiment on altruism, empathy, and charitable giving*. Nber Working Paper Series (Working Paper 17648).
- Baran, N., Sapienza, P., & Zingales, L. (2009). *Can we infer social preferences from the lab? Evidence from the trust game*. NBER Working Paper Series (Working Paper 15654).
- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, *11*(2), 122–133.
- Barr, A., & Zeitlin, A. (2010). *Dictator games in the lab and in nature: external validity tested and investigated in Ugandan primary schools*. CSAE Working Paper.
- Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—Evidence from donations. *Experimental Economics*, *11*(3), 268–283.
- Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(3), 321–338.
- Bridges, F. S., Thompson, P. C., & Coady, N. P. (2001). Homeless affiliation, location, and sender address in return of lost letters. *Journal of Social Distress and the Homeless*, *10*(3), 235–242.
- Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction*. New York: University Presses CA.
- Camerer, C. F. (2011). *The promise and success of lab-field generalizability in experimental economics: a critical reply to Levitt and List*. SSRN Working Paper.
- Card, D., DellaVigna, S., & Malmendier, U. (2011). The role of theory in field experiments. *Journal of Economic Perspectives*, *25*(3), 39–62.
- Carpenter, J., & Meyers, C. K. (2010). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics*, *94*, 911–920.
- Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, *92*(4), 1218–1221.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354.
- Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610.
- Englmaier, F., & Gebhardt, G. (2010). *Free riding in the lab and in the field*. Governance and the efficiency of social systems (GESY), CESifo Working Paper No. 344.
- Fehr, E., & Leibbrandt, A. (2008). *Cooperativeness and impatience in the tragedy of the commons*. Institute for Empirical Research in Economics, University of Zurich Working Paper No. 378.
- Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, *81*(1), 74–81.
- Frey, B. S., & Meier, S. (2004). Pro-social behavior in a natural setting. *Journal of Economic Behavior & Organisation*, *54*(1), 65–88.
- Gurven, M., & Winking, J. (2008). Collective action in action: prosocial behavior in and out of the laboratory. *American Anthropologist*, *110*(2), 179–190.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *43*(4), 1009–1055.
- Hoffman, E., McCabe, K., & Smith, V. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, *86*(3), 653–660.
- Howitt, D., & McCabe, J. (1978). Attitudes do predict behaviour—in mails at least. *British Journal of Social and Clinical Psychology*, *17*(3), 285–286.
- Howitt, D., Craven, G., Iveson, C., Kremer, J., McCabe, J., & Rolph, T. (1977). The misdirected letter. *The British Journal of Social and Clinical Psychology*, *16*, 285–286.
- Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, *95*(5), 1688–1699.
- Kremer, J., Barry, R., & McNally, A. (1986). The misdirected letter and the quasi-questionnaire: unobtrusive measures of prejudice in Northern Ireland. *Journal of Applied Social Psychology*, *16*(4), 303–309.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153–174.
- Levitt, S. D., & List, J. A. (2009). Field experiments in economics: the past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482–493.
- List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, 25(3), 3–16.
- Merritt, C. B., & Fowler, R. G. (1948). The pecuniary honesty of the public at large. *Journal of Abnormal and Social Psychology*, 43(1), 90–93.
- Milgram, S. (1969). The lost letter technique. *Psychology Today*, 3(1), 30–33, 66–68.
- Milgram, S., Mann, L., & Harter, S. (1965). The lost letter-technique: a tool of social research. *Public Opinion Quarterly*, 29(3), 437–438.
- Sechrest, L., & Belew, J. (1983). Nonreactive measures of social attitudes. In L. Bickman (Ed.), *Applied social psychology annual* (Vol. 4, pp. 23–64). Beverly Hills: Sage.
- Stoop, J. (2012). *From the lab to the field: envelopes, dictators and manners*. Munich Personal RePEc Archive, MPRA Paper No. 37048.
- Wilson, D. S., O'Brien, D. T., & Sesma, A. (2009). Human prosociality from an evolutionary perspective: variation and correlations at a city-wide scale. *Evolution and Human Behavior*, 30(3), 190–200.