On the origins of human sociality

Akdeniz, A.

Publication date
2023

Chapter 3

The evolution of honesty by partner choice[1]

3.1 Introduction

Why people lie is no mystery. Everyone acts on the basis of what they think is true, and if one can manipulate what others believe, then one can manipulate what others do. Why honesty exists, on the other hand, is not immediately clear. Revealing information because it is true, even if doing so would have negative consequences for oneself, requires some mechanism that compensates for the cost of honesty. Explanations that have been suggested for the evolution of honesty are similar to the ones that are proposed for the evolution of cooperation or altruism. Some authors suggest lying aversion can be explained by population structure (Krakauer and Pagel, 1995); others suggest that it can be explained by interactions being repeated (Rich and Zollman, 2016). We explore the possibility that, while telling the truth can be costly, being committed to telling the truth can be beneficial, because it makes one a more attractive partner.[2]

In situations in which one party knows or observes something that the other does not, honesty can make a difference since being honest can help commit to prosocial behaviour. In a partnership where only one of the partners for instance observes the result of a joint effort, the partner that is in the know would have an incentive to misinform the other

[1]This chapter is based on joint work with Stephan Jagau, Shaul Shalvi, and Matthijs van Veelen.

[2]Partner choice (Heintz et al., 2016) and commitment (Akdeniz and van Veelen, 2021; Frank, 1987; 1988) have been suggested as explanations for cooperative or pro-social behaviours in general, but here we focus on honesty (notice that in Heintz et al. (2016), honesty is an umbrella term for anything prosocial, and does not mean lying aversion). In the Discussion section we will come back to the relation between commitment and partner choice, and briefly discuss how it is also possible to have one without the other.

parties, and underreport the true gains in order to appropriate a larger portion of it for herself. When choosing a partner, one would therefore be happy to find someone who cannot help but tell the truth, and thereby opens herself up to scrutiny by the other parties. This preference for honest partners in mutually beneficial exchanges could then balance against the costs of honesty, and sustain a preference for truth-telling in human societies.

In this paper, we investigate the mechanism outlined above. We first of all test if, in a game with asymmetric information, honest others are indeed preferred as partners, and if honest partners indeed behave more prosocially. We find that the answer to both of these questions is yes. The findings from our lab experiment show that, in a trust game where only one of the partners observes the multiplier, honest individuals behave on average more prosocially than dishonest individuals, and that this is anticipated by others and creates a preference for honest partners. We also ask the question if the pathway for the connection between honesty and prosociality is conscious, in the following sense. Lying averse people by definition have a hard time lying. Therefore, when doing something not particularly prosocial is accompanied by a choice between, on the one hand, lying and, on the other, not lying and revealing they did something selfish, they could prefer to do the prosocial thing in order to avoid putting themselves in this difficult position. We find no evidence for this pathway. Instead, the outcome of our experiment is consistent with the possibility that lying averse people might have a harder time justifying selfish behaviour to themselves in the first place, or that they are just across the board more prosocial people.

3.2 Experimental design

We use a two-part design to examine the role of honesty in partner choice in a situation with asymmetric information. In the first part, we measure the relative lying aversion of subjects using a die-rolling task (Bicchieri et al., 2020; Conrads et al., 2013; Fischbacher and Föllmi-Heusi, 2013; Leib et al., 2021; Maréchal et al., 2017; Shalvi et al., 2011). Subjects observe a die-roll on their screen and are asked to report the outcome. Which die-roll a subject observes is chosen randomly, and each roll is equally likely. This happens for three rounds. Every time a subject reports a 5, she earns 75 points; and every time she reports another number, she earns 0 points. The potential earnings in the die-rolling task are therefore 0, 75, 150, or 225 points, depending on how many times a subject reports a 5. We define the lying aversion of a subject as the ratio of times she did not misreport a 5 out of the number of times she could have misreported a 5.

In the second part of the experiment, we assign subjects to groups of three. The group assignment within a session is done as follows. We first randomly assign one third of the subjects to the role of trustor. The remaining subjects are assigned the role of trustee. Each trustor is then matched with two trustees, one randomly drawn from the more lying averse half of the trustee population, and the other randomly drawn from the less lying averse half. In these groups, they play a version of the trust game for one round (Berg et al., 1995; Clots-Figueras et al., 2016; Güth et al., 2014), where trustors do not choose how much to send, but who to send it to, and where the multiplier is uncertain. Each group consists of one trustor, who is endowed with 50 points, and two candidate trustees, who are endowed with 0 points. The trustor observes the die-rolls that the candidate trustees saw, as well as what they reported, and chooses a partner to send her endowment to. The points sent by the trustor are then multiplied, either by 2 or by 4. Which one it is, is determined by chance, and both multipliers are equally likely. The trustees observe the multiplier, but the trustor does not. The trustees then choose how much they want to send back to the trustor, if they are chosen. While trustees make their decisions, we also elicit how much the trustor expects the trustees to send back, in case they are chosen, and depending on the multiplier.

We have two, between-subjects treatments: communication (C) and no communication (NC). In the communication treatment, there is cheap-talk messaging, where the candidate trustees also send a message about the multiplier to the trustor, together with the amount of points to send back. They choose between "the multiplier is 2" and "the multiplier is 4". Hence, if the multiplier is low, there is not much room for trustees to adjust their behaviour; however, if it is high, trustees face a choice between behaving prosocially, or increasing their payoff without damaging their image by pretenting that the multiplier is low. In the no communication treatment, there is no messaging.

The two treatments are there in order to investigate the mechanism by which honest subjects end up behaving more prosocially. Not all people are selfish, but even selfish people generally are not too keen on admitting they are selfish, and this is true also in experiments in which there is no way in which gaining a reputation for selfishness can have repercussions (Andreoni and Bernheim, 2009; Dana et al., 2007; Pillutla and Murnighan, 1995). Therefore, we assume that subjects in our experiment also generally care about not appearing selfish in their trust game behaviour. Assuming that they do, subjects that are not lying averse can have their cake and eat it –to some degree, if the multiplier is 4. If it is, they can report that the multiplier is 2, and return an amount that would be fair if the multiplier was 2, but not if the multiplier is 4. This way they seem prosocial, in so far as the trustor believes the message, but keep more money for themselves. For lying averse subjects, this would be harder. When the multiplier is 4, the choice for lying averse subjects is likely to become one between, on the one hand, acting prosocially, and on the other hand, acting selfishly and admitting to it. In that case, they might just choose to always return a fair amount for the true multiplier, also if it is high, and tell the truth. If this is the mechanism by which lying averse individuals end up making more generous choices, then a difference between the more and less lying averse would occur in the communication treatment, and only if the multiplier is 4, that is, when there is something to lie about. In the no communication treatment, trustors may form a belief about the multiplier based on the amount that the trustee sends back, but in the absence of the possibility to make any explicit statements about the true multiplier, we would expect the impact of lying aversion on prosociality to be smaller or non-existent.

Once the candidate trustees decide how much they would like to send back if they are chosen (and their messages in the communication treatment), and the trustor submits her beliefs, the trust game ends. The trustor earns the points sent back by the trustee she chose as her partner, and she earns additional points from the belief elicitation. In the communication treatment, the trustor also observes the message sent by the trustee she chose, and responds by choosing between "I believe you" and "I do not believe you". The chosen trustee earns the amount she received (50 points sent by the trustor times the multiplier) minus the amount she chose to send back. In the communication treatment, she also observes the response of the trustor to her message about the multiplier. The non-chosen trustee earns the outcome of a random draw between 0 and 100, if the multiplier is 2; and between 0 and 200, if the multiplier is 4. The reason behind the random draw for the trustee that is not chosen, and the relatively large amounts that can be earned in the first part, is that we do not want subjects to be strategically honest in the die-rolling task. Being chosen in the second part is in expectation still better than not being chosen, but the difference is small enough to make sure that, if they are in the lab just to make money, the best way to do that is to lie in the first part. This may seem a counter-intuitive choice, but the purpose of this experiment is not to show that honest people can walk out of the lab with more money than dishonest people. We test hypotheses about how the benefits of being committed to telling the truth come about – and a defining characteristic of being committed to telling the truth, is that one does not immediately switch to lying, whenever that is more opportune.

After the trust game, we elicit participants' social value orientation (SVO), and collect some demographic information such as gender and age (see the Methods section in the

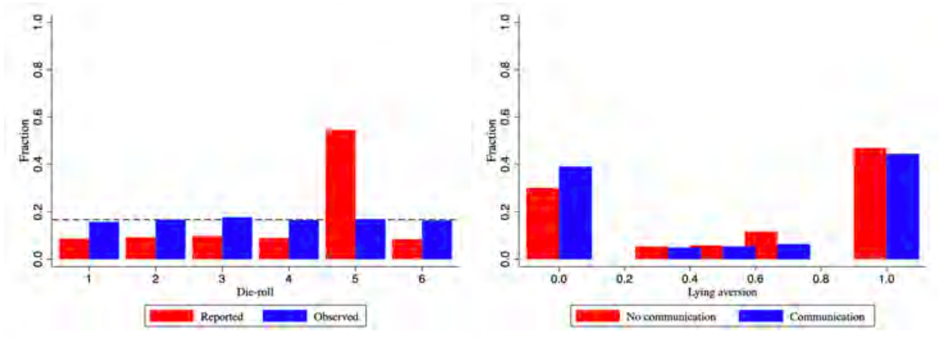
Supplementary Information for details).

If honesty evolved due to its hypothesized role in partner choice, we should first of all observe that honest individuals act more prosocially. Our first hypothesis is therefore that honest trustees would send back more points compared to dishonest trustees in the trust game. Moreover, if the reason why honest partners act more prosocially is the choice between, on the one hand, admitting to selfish behaviour and, on the other hand, behaving prosocially and telling the truth, we should observe that a stronger connection between honesty and prosociality in the presence of communication. Our second and third hypotheses are therefore that the average shares of points sent back and the difference between the shares of points sent back by honest and dishonest trustees would be larger in the communication treatment than in the no communication treatment, respectively. We expect both of these effects to be driven by the cases where the multiplier is high, since only then there is a reason to deceive one's partner. The choosing side should moreover anticipate the benefits of, and have a preference for, honest partners, for honesty to play a role in partner choice. Therefore, our fourth hypothesis is that trustors would expect honest trustees to send back more points, and choose them more often as their partners than dishonest trustees. Finally, we hypothesize that trustors would anticipate an impact of communication, especially on honest trustees (which are our fifth and sixth hypotheses; see the Methods section in the Supplementary Information for details on our hypotheses).

3.3 Results

Below we will first summarize the main findings from the die-rolling task and the trust game, and then give a detailed account of the results on trustee behaviour, trustor expectations and partner choice in the trust game. In our analyses, we normalize how much trustees send back and trustors expect to receive back by considering the proportion of the total amount available to the trustee that she returns and is expected to return.

Figure 3.2.1 presents the results of the die-rolling task. The distribution of reported rolls differ significantly from the distribution of actual rolls, with a spike at the reports of 5, and a corresponding downward shift on the reports of other numbers (Figure 3.2.1a). The majority of subjects either never lie, even though they could have benefited from lying, or they lie every time they can benefit from it, with no significant difference between the two treatments (Figure 3.2.1b). Since our experiment is designed in a way so that the groups in the trust game contain one trustee from the more lying averse half of the population

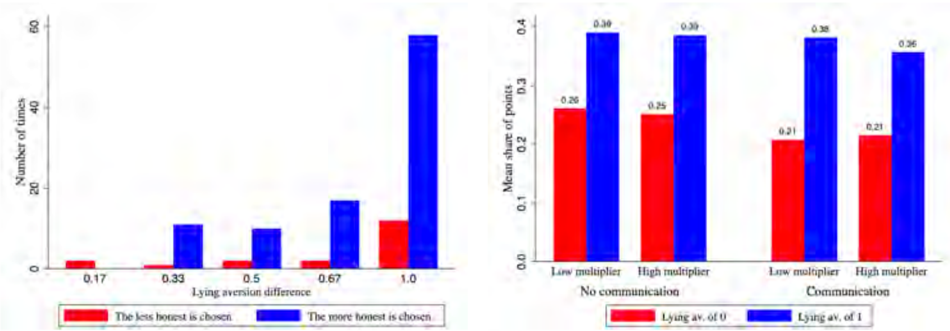


(a) Observed and reported die-rolls **(b) Distribution of lying aversion**

Figure 3.2.1: Lying aversion. (a) The frequencies of die-rolls observed by subjects (blue bars) is close to 1/6 for every number between 1 and 6, as expected. The frequencies of reported die-rolls (red bars) are consistent with significant over-reporting of 5s, but also considerable truthful reporting. **(b)** Our measure of lying aversion is the ratio of the number of instances in which the subject could have benefited from lying, but chose not to, and the number of instances they could have benefited from lying. For someone that observed zero fives, this number therefore can be 0, 1/3. 2/3 or 1 – where 0 means the subject always lied and 1 means the subject never lied. For someone that observed one five, this number can be 0, 1/2 or 1, for someone that observed 2 fives, 0 or 1, and for the 2 subjects that observed 3 fives, this is not defined (who are therefore excluded from our analyses.

of trustees and one trustee from the less lying averse half within a session, this typically generates relatively diverse pairs (see also Figure 3.3.1a). In the remainder of the paper, we will restrict our analyses on trustee behaviour and trustor expectations to trustees with a lying aversion measure of 1 or 0 to simplify the classification of individuals as honest and dishonest. Our results are however robust to the inclusion of all trustees (see the Supplementary Information for the analyses on the full sample).

If we look at the data for both treatments together, we first of all find that trustors mostly choose the more honest trustee as their partner (Figure 3.3.1a). They also expect honest trustees to send back more points (Figure 3.3.1b); and they are right in doing so, although they do overestimate the difference between honest and dishonest trustees (Figure 3.3.2). Next we look at the treatment effect to examine whether honest people act prosocially to avoid lying to their partners. We find no evidence for this. If we consider the behaviour when the multiplier is high in the communication treatment, we do not find a significant difference between subjects with a lying aversion measure of 0 and subjects with a lying aversion measure of 1. Instead, we find quite the opposite, as the overall difference between the behaviour of subjects with a lying aversion measure of 0 and subjects with a lying aversion measure of 1 seems to be driven mostly by the difference in



(a) Partner choice **(b) Trustor expectations**

Figure 3.3.1: Trustor behaviour. **(a)** Partner choice decisions of trustors. In red the number of times the less honest candidate is chosen, in blue the number of times the more honest partner is chosen, both depending on the difference in lying aversion of the two candidate trustees. **(b)** Averages of the shares of points trustors expected to be returned, based on the lying aversion measure of the trustees, the treatment and the multiplier in the trust game.

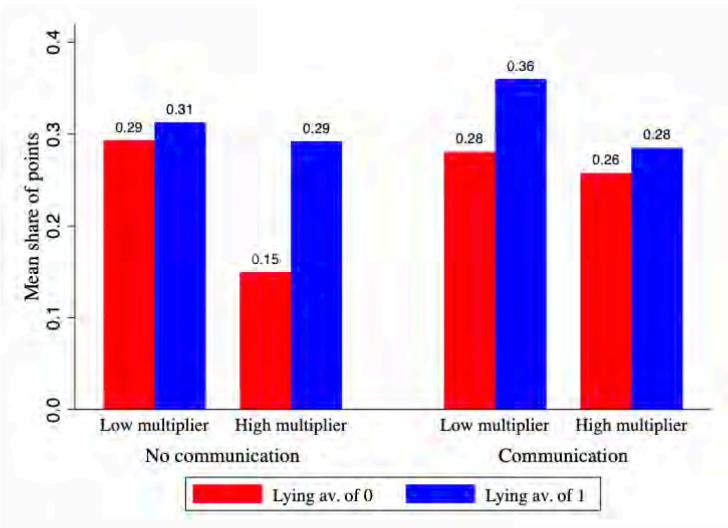


Figure 3.3.2: Trustee behaviour. Averages of the shares of points sent back by trustees, based on the lying aversion measure of the trustees, the treatment and the multiplier in the trust game.

the no communication treatment, when the multiplier is high (see Figure 3.3.2). Next we discuss trustee behaviour, trustor expectations and partner choice in more detail.

Trustee behaviour We can start with a simple statistical analysis, where we just compare

NC & Low multiplier	0.9706
NC & High multiplier	0.0018
C & Low multiplier	0.2391
C & High multiplier	0.6828
Low multiplier	0.4168
High multiplier	0.0169
NC	0.0913
C	0.2429
Overall	0.0413

Table 3.3.1: Mann Whitney test results. p-values of the Mann Whitney tests comparing the returns by trustees with a lying aversion measure of 0 versus 1, based on the treatment and/or the multiplier. The last row reports the result for the overall comparison where we aggregate observations for both treatment and multiplier. In the cases where there is a significant difference, those with a lying aversion of 1 send back more points than those with a lying aversion of 0.

the shares of points returned by those with a lying aversion measure of 0 with that of those with a lying aversion measure of 1 for different partitions of the sample (Table 3.3.1). If we do that for the combinations of treatment and multiplier depicted in Figure 3.3.2, then we only see a significant difference between the not lying averse and the lying averse for the high multiplier in the NC treatment (p-value: 0.0018, which rejects the null hypothesis of equality against the Bonferroni corrected critical value for multiple hypothesis testing). None of the other differences are anywhere close to significant (p-values: 0.9706 for the low multiplier in NC, 0.2391 for the low multiplier in C, and 0.6828 for the high multiplier in C). Here we should stress that we hypothesized that there would be a larger difference between the lying averse and the not lying averse for the combination of high multiplier and communication, than for the combination of high multiplier and absence of communication. What we observe however is that the difference is non-existent in the first case, whereas it is striking in the latter. A remarkable feature of the data is that the average shares that trustees return are not too far away from 30% in all cases, with one exception. That one exception is that when the multiplier is high and there is no communication, then those that are classified as not lying averse on average return 15% (Figure 3.3.2). This is remarkable, because, obviously, 15% of the amount for a multiplier of 4 equals 30% of the amount for a multiplier of 2.

If we aggregate observations for both multipliers, and compare the returns by honest and dishonest trustees within each treatment separately, then neither of the two is significant (p-values: 0.0913 for NC, and 0.2429 for C). If, on the other hand, we aggregate observations for both treatments, and compare the returns by honest and dishonest trustees for each multiplier separately, again neither of the two is significant (p-value: 0.0169 for the high multiplier, p-value: 0.4168 for the low multiplier, compared to the Bonferroni corrected critical value for multiple hypothesis testing). If we aggregate observations for both multipliers and treatments, the difference between the shares of points sent back by the lying averse and the not lying averse is only marginally significant (p-value: 0.0413).[3] In the cases where there is a significant difference between the two groups, those with a lying aversion of 1 are the ones that act more prosocially.

Also a more elaborate statistical analysis confirms this. Table 3.3.2 presents ordinary least squares (OLS) regression results, where the dependent variable is the the share of points sent back by the trustee. The independent variables are the lying aversion of the trustee (for testing our first hypothesis), the level of the multiplier, the communication treatment, the interaction variable between the treatment and the multiplier (for testing our second hypothesis), and the interaction variable between the treatment, the multiplier and the lying aversion of the trustee (for testing our third hypothesis). Results show that honest trustees behave more prosocially than dishonest trustees in the no communication treatment (Column 1) and overall (Columns 3-5), but not in the communication treatment (Column 2). Altogether, these findings provide support for our first hypothesis that honest individuals act more prosocially than dishonest individuals in a situation with asymmetric information.

We had also hypothesized that a pathway from honesty to prosociality would be that lying averse individuals might prefer to send back a fair amount of points corresponding to the true multiplier, in order to avoid the choice between lying and revealing having been selfish. When the multiplier is low, this does not create a problem, because one can then only send back a decent amount for the low multiplier. When there is no communication, this also would not create a problem, because they are not asked to report the multiplier. Therefore, we expect this effect to be present when the multiplier is high, and where there is communication. In line with Figure 3.3.2 and the results of the non-parametric tests in Table 3.3.1, this is not what we find, as shown by the insignificance of the interaction variables in Columns 4-6 of Table 3.3.2. This provides evidence against our second and third hypotheses that there would be higher levels of prosociality in the presence of communication, due to honest individuals acting more prosocially to avoid admitting to selfish behaviour.

[3]We can run a robustness check on this result using the SVO measures collected at the end of the experiment. Mann-Whitney test comparing the SVO angles of trustees with a lying aversion measure of 0 versus 1 suggests a statistically significant difference between the two groups of trustees (p-value: 0.0037).

	(1) Ratio	(2) Ratio	(3) Ratio	(4) Ratio	(5) Ratio	(6) Ratio	(7) Ratio
Lying av.	0.0781**	0.0537	0.0649**	0.0657**	0.0786**	0.0422	
	(0.032)	(0.161)	(0.013)	(0.013)	(0.015)	(0.173)	
Multiplier	-0.0735**	-0.0512	-0.0613**	-0.0718**	-0.0736**	-0.0321	
	(0.040)	(0.173)	(0.017)	(0.041)	(0.038)	(0.430)	
C			0.0282	0.0179	0.0172	0.0192	
			(0.268)	(0.661)	(0.675)	(0.640)	
(C=1)#(M=1)				0.0209	0.0500	-0.0194	
				(0.684)	(0.425)	(0.729)	
(C=1)#(M=1)#(LA=1)					-0.0511		
					(0.335)		
(C=0)#(M=1)#(LA=0)						-0.100*	-0.149***
						(0.072)	(<0.001)
Constant	0.264***	0.295***	0.265***	0.270***	0.264***	0.282***	0.299***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N	111	111	222	222	222	222	222
Treatment	NC	C	Both	Both	Both	Both	Both
adj. R^2	0.053	0.018	0.042	0.038	0.037	0.046	0.045

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3.2: OLS regression results on trustee behaviour. Share of points sent back by trustees, based on the lying aversion measure of the trustee; in the NC treatment (Column 1), in the C treatment (Column 2), in both treatments including the treatment variable (Column 3), the interaction variable between the treatment and the multiplier (Column 4), and the interaction variable between the treatment, the multiplier, and the lying aversion of the trustee (Column 5). Columns 6 and 7 report the results for our unexpected finding, where the interaction variable is included for the NC treatment when the multiplier is high for the not lying averse subjects. The data is restricted to the trustees with a lying aversion measure of 0 or 1. For the interaction variables, C=1 (0) means (no) communication treatment, M=1 means high multiplier, and LA=1 (0) means lying aversion measure of 1 (0).

Instead of the difference being driven by the lying averse sending back larger shares when the multiplier is high and there is communication, also with the OLS regression results, we observe that it is mainly driven by them sending back more when the multiplier is high *and there is no communication* (see Columns 6-7 in Table 3.3.2). Our communication treatment therefore appears to have a prosocializing effect on behaviour, however, in contrast to our expectations, it is on the not lying averse individuals. One interesting

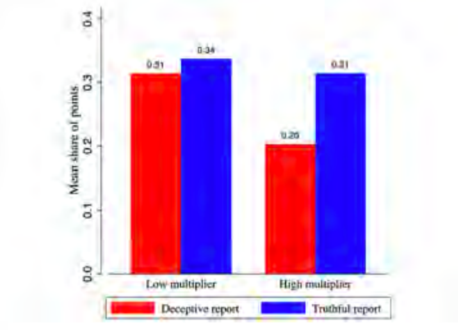


Figure 3.3.3: Trustee behaviour based on misreporting in the communication treatment. Averages of the shares of points returned by trustees based on whether the trustee misreported the multiplier in the communication treatment. When the multiplier is high, those who report truthfully send back on average 31% of the available points, and those who misreport send back on average 20%. The difference is statistically significant.

point here is that relatively few trustees make use of the communication opportunity to keep a prosocial image while being selfish, and that those who make use of it are not necessarily the ones classified as dishonest in the die-rolling task. Out of the 55 trustees facing the high multiplier in the communication treatment, 19 misreport the multiplier. Those who do so act less prosocially compared to those who report truthfully (Mann-Whitney test p-value: 0.008, see also Figure 3.3.3). There however does not seem to be a correlation between misreporting in the trust game and misreporting in the die-rolling task. If we regress the probability of misreporting in the trust game on the lying aversion of the trustee from the die-rolling task, the estimated coefficient on lying aversion is highly statistically insignificant (p-value: 0.992, see the Supplementary Information for details). The disconnect between misreporting in the two parts of the experiment could however be due to the relatively small number of observations of misreports in the second part of our experiment. Out of the 56 trustees facing the low multiplier in the communication treatment, 7 misreport the multiplier. Those who do so, however, still send back amounts that are consistent with the low multiplier (see the left-most bar in Figure 3.3.3). This, in combination with the low number of observations in this case, suggests that reporting a high multiplier, when it is in fact low, is likely to be a mistake.

Trustor expectations Table 3.3.3 presents the OLS regression results, where the dependent variable is the share of points expected back by the trustor. The independent variables are the lying aversion of the trustee (for testing our fourth hypothesis), the level of the multiplier, the communication treatment, the interaction variable between the

	(1) Ratio	(2) Ratio	(3) Ratio	(4) Ratio	(5) Ratio	(6) Ratio
Lying av.	0.131***	0.158***	0.145***	0.145***	0.146***	0.143***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Multiplier	-0.00676	-0.0100	-0.00842	-0.00676	-0.00676	-0.00676
	(0.377)	(0.352)	(0.201)	(0.373)	(0.374)	(0.375)
C			-0.0303	-0.0286	-0.0286	-0.0249
			(0.275)	(0.318)	(0.321)	(0.396)
(C=1)#(M=1)				-0.00329	-0.00101	0.000149
				(0.802)	(0.961)	(0.994)
(C=1)#(M=1)#(LA=1)					-0.00421	-0.00635
					(0.888)	(0.832)
Lying av. trustor						0.0427
						(0.181)
Constant	0.260***	0.216***	0.253***	0.252***	0.251***	0.225***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N	216	222	438	438	438	438
Treatment	NC	C	Both	Both	Both	Both
adj. R^2	0.089	0.210	0.146	0.144	0.142	0.149

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3.3: OLS regression results on trustor expectations. Share of points expected back by trustors, based on the lying aversion measure of the trustee; in the NC treatment (Column 1), in the C treatment (Column 2), in both treatments including the treatment variable (Column 3), the interaction variable between the treatment and the (hypothetical) multiplier (Column 4), the interaction variable between the treatment, the (hypothetical) multiplier, and the lying aversion of the trustee (Column 5), and the lying aversion of the trustor (Column 6); with the errors clustered at the trustor level since each trustor submits multiple beliefs. The data is restricted to trustor expectations for the trustees with a lying aversion measure of 0 or 1. For the interaction variables, C=1 means communication treatment, M=1 means high multiplier, and LA=1 means lying aversion measure of 1 for the trustee.

treatment and the multiplier (for testing our fifth hypothesis), and the interaction variable between the treatment, the multiplier, and the lying aversion of the trustee (for testing our sixth hypothesis). Results show a very strong positive correlation between how lying averse one's partner is and how prosocial one expects them to be (Columns 1-6). Trustors expect to receive on average 37.9% of the available points from trustees with a lying aversion measure of 1, while they expect only 23.3% from trustees with a lying aversion

	(1) Chosen	(2) Chosen	(3) Chosen	(4) Chosen
The more lying averse trustee	1.576***	1.540***	1.620***	1.653***
	(<0.001)	(0.001)	(<0.001)	(<0.001)
C		0.0690		-0.0637
		(0.914)		(0.900)
N	70	70	115	115

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3.4: Logit regression results on partner choice. The logistic regression for the probability that the more lying averse candidate trustee is chosen as a partner by the trustor. Columns 1 and 2 restrict the data to the trustee pairs with a lying aversion difference of 1, including and excluding the treatment variable; and Columns 3 and 4 include the data on all trustee pairs with a non-zero lying aversion difference, including and excluding the treatment variable.

of 0 (Mann-Whitney test p-value < 0.001). They do not anticipate any impact of the communication treatment (Columns 3-5). These findings provide support for our fourth hypothesis on trustors anticipating honest individuals to be more prosocial, but against our fifth and sixth hypotheses on the anticipation of an impact of communication on the trustee population as a whole, or on the lying averse trustees.

The finding that trustors place great value on the information about the trustees' lying aversion, but do not anticipate any other effect, might be explained by the general pattern in trustee behaviour. As shown previously, honest and dishonest trustees behave similarly in the communication treatment, and in the no communication treatment when the multiplier is low; and honest trustees behave significantly more prosocially in the no communication treatment when the multiplier is high. Therefore, although it might not always benefit the trustor to pick an honest partner, it never hurts. We expect this to be a general pattern, also outside of the lab. Picking an honest partner is likely to be beneficial, but unlikely to be harmful in many partnerships. The selective pressure on the choosing side could therefore be focused mostly on the detection of honesty, and not so much on how the specifics of an interaction affect the behaviour of honest versus dishonest partners. While gathering information likely involves cognitive and other costs (Drugowitsch et al., 2012; Sullivan, 1994), it might not bring additional benefits, once we know how lying averse our partner is. This could have made the choosing side more attentive to information on lying aversion compared to other, seemingly less important, details of an interaction.

Partner choice Table 3.3.4 presents the logistic (logit) regression results, where the dependent variable is the probability that the more honest trustee in a trustee pair is chosen. Columns 1 and 2 restrict the data to trustors who were presented with a trustee pair with a lying aversion difference of 1, and Columns 3 and 4 include data for all trustors who were presented with a trustee pair with any non-zero lying aversion difference. Results show that honest trustees are chosen significantly more often as partner, with no significant difference between treatments (Columns 2 and 4). In line with the previous findings on the share of points trustors expect from honest versus dishonest trustees, honest trustees are therefore trusted significantly more often in a situation with asymmetric information. These findings provide evidence for our last hypothesis, which concerns the existence of a preference for honest partners.

One prerequisite for honesty to play a part in partner choice is obviously that others can tell. Trustors in our experiment make their partner choice decisions with full information on the lying aversion of the candidate trustees –as finely measured as possible by the die-rolling task. To what extent we are able to tell who is honest and who is not in real life, is an interesting open question. If possible, one would like to be perceived as honest, but lie when opportune, and an invasion of a type that seems honest, but is not, would undermine the selective advantage of the truly honest. There would therefore be selection pressure also on the ability to detect true commitment to honesty. Even though deception occurs regularly in real life, and sometimes with substantial consequences, there is considerable evidence suggesting that one can tell whether someone is (being) honest under certain conditions. For instance, people seem to be relatively good at detecting deception when they know their interaction partners (Anderson et al., 2002; DePaulo, 1994; DePaulo and Kashy, 1998; Ten Brinke et al., 2016; Von Hippel and Trivers, 2011; Vrij and Mann, 2005). Also, our brains are argued to have specialized modules to detect cheating (Cosmides and Tooby, 1992; Stone et al., 2002). The fact that honesty still is one of the most important virtues one looks for in prospective partners (Cottrell et al., 2007; Regan et al., 2000) moreover suggests that our detection capabilities are adequate enough to be utilized in judging others.[4]

3.4 Discussion

Why honesty exists is an open question. Truthfully revealing information, that could otherwise be used for one's benefit, requires a mechanism that compensates for the costs

[4]See also Heintz et al. (2016) for a discussion on how social traits and strategic vigilance, that is the ability to spot genuine displays of those traits, could have co-evolved.

of honesty. We investigate how partner choice can bring such benefits and enable the evolution of honesty. More specifically, we study whether honesty credibly signals commitment to prosociality, and create a demand for honest partners in mutually beneficial interactions.

Most partnerships involve reporting private information that individuals have. Since the interests of the involved parties do not usually align, the partner that is in the know would be tempted to deceive others for her own benefit. This temptation would likely destabilize partnerships unless partners can find a way of committing to treating each other fairly. This is where honesty can offer a solution. By increasing the (cognitive) costs of lying, it can make one prefer behaving prosocially to avoid lying to others, and enable trust between partners. The choice between behaving prosocially and lying about having been selfish is especially pertinent when partners are explicitly communicating with each other. When there is no communication between partners, the commitment power of honesty to prosocial behaviour would no longer be relevant.

Using a lab experiment, we test the mechanism outlined above. We examine whether honest people are in general more prosocial in a situation with asymmetric information, whether the connection between honesty and prosociality goes through the impact of communication, and whether these are anticipated by others and create a preference for honest partners. The results of our experiment show that honest people are indeed on average more prosocial in a trust game with an uncertain multiplier, that this is anticipated by others and results in a preference for honest partners in the trust game. However, as indicated by the null impact of our communication treatment on honest trustees, the connection between honesty and prosociality does not seem to go through communication. Honest individuals behave consistently prosocially independent of whether there is communication or not. Instead, dishonest individuals react to our communication treatment. They behave significantly less prosocially than honest individuals when this requires them to deceive their partners only implicitly (in the no communication treatment), but not when it requires an explicit lie to their partners (in the communication treatment). Below we discuss potential reasons for our findings on communication.

Deception in real life takes many different forms. People lie by both commission and omission, and by behaving in a way that would lead their partners to hold inaccurate beliefs. In our no communication treatment, trustees can lead their partner to wrongfully believe that the multiplier is low by sending back an amount that would be consistent with the low multiplier, when it is high. How much one sends back therefore has informational value by itself. When deception involves such behavioral misleading and not explicit misreporting, we observe that individuals classified as dishonest, but not as honest, deceive their partners by sending back amounts that would be consistent with the low multiplier. Honest trustees however send back amounts consistent with the actual multiplier they observe. This suggests that our die-rolling task could be capturing an aversion to both implicit behavioural misleading and explicit verbal misleading, where those classified as dishonest in the die-rolling task still avoid the latter.

Both honest and dishonest subjects in our experiment generally refrain from lying in the trust game. The existence of a partner on the receiving end of a lie, without any strategic role of communication, therefore seems to push the majority of our subjects to report the multiplier truthfully (and behave accordingly prosocially). Moreover, in contrast to the findings on the ultimatum game (Boles et al., 2000; Kriss et al., 2013), introducing the possibility to send a deceptive message about the size of the pie increases prosocial behaviour in our experiment. The decision to *trust* in our trust game, as opposed to no such decision in the ultimatum game, might therefore have created an interaction environment that is less prone to deception. The timing of communication could have further reinforced this prosocializing impact of communication in our experiment. Existing literature suggests that the timing of communication affects one's propensity to lie and that people are most prosocial closest to the time of communication (Andersson and Wengström, 2012; Bhattacharya et al., 2020; Blume and Ortmann, 2007; Casella et al., 2018; Gneezy, 2005). Since communication happens at the same time as the choice of prosociality in our experiment, it might have nudged not only the honest but also the dishonest to report truthfully. It is therefore important in future research to investigate the impact of other forms of communication on prosociality of honest individuals, especially in settings where communication can be used to influence the recipient's behaviour (Blume et al., 1998; Cai and Wang, 2006; Crawford and Sobel, 1982; Sánchez-Pagés and Vorsatz, 2007).

Here we examined how honesty can be advantageous in partner choice. However, there is another, additional channel through which the ability to commit can help an individual. Honesty can bring benefits not only by increasing the chances that one is chosen as a partner, but also by changing how one's partner behaves during a partnership. Consider, for instance, the standard trust game (Berg et al., 1995; Clots-Figueras et al., 2016; Güth et al., 2014), where the trustor does not choose who to partner with, but chooses how much of her endowment to send to her existing partner. The partner choice channel is therefore absent in this version, but the trustee's honesty can still change the trustor's behaviour to her benefit. Given that honest individuals act on average more prosocially than dishonest ones, and that this is anticipated by others, trustors are likely to be willing to send larger amounts to honest trustees. This would increase the size of the pie to be shared, and result in larger gains to be attained from the partnership. Partner choice is therefore one of the channels through which commitment power of honesty to prosociality can work, but it is not the only one; the ability to influence partner behaviour in a given interaction brings benefits as well (see Akdeniz and van Veelen (2021) for an overview on the role of commitment in a broad spectrum of human social behaviours).

Appendix

3A.1 Methods

General outline and procedure of the experiment[5] The experiment is conducted at the CREED lab of the University of Amsterdam. It is computerized using oTree (Chen et al., 2016). We invited 411 subjects to the lab, with 207 (204) subjects randomly assigned to the NC (C) treatment. We ran three pilot sessions with a total of 57 subjects before running our main sessions. After the pilot sessions, we updated our instructions and added comprehension questions. The data from the pilot sessions are excluded in our analyses. All participants received their earnings from the experiment plus a €4 show-up fee. Experimental payoffs are expressed in points during the experiment. At the end of the experiment the points are converted into euros at a conversion rate of €0.05 per point. The experiment consists of four parts and a survey. In the first part of the experiment, subjects are asked to report the outcome of a die roll they see on their screen for three rounds (Conrads et al., 2013; Fischbacher and Föllmi-Heusi, 2013; Leib et al., 2021; Maréchal et al., 2017; Shalvi et al., 2011). In the second part, subjects are assigned into groups of three to play a simplified version of the trust game (Berg et al., 1995) with partner choice and an uncertain multiplier. In the third and fourth parts, subjects make six resource allocations decisions (Crosetto et al., 2019; Murphy et al., 2011) and four investment decisions (Gneezy and Potters, 1997),[6] respectively. At the end of the experiment, subjects are asked to complete a survey. Subjects are paid their earnings from either the first part or the second part of the experiment, in order to eliminate potential income effects in the second part (Drouvelis and Sonnemans, 2017). Which stage is paid is determined randomly. Additionally, they are paid for two of the resource allocation tasks – once in the sender and once in the receiver role, and for one round of the investment task. Which of these tasks are paid is also determined randomly. The participation fee and any additional amount of money participants earned is paid to them privately in cash at the end of the experiment. Subjects were matched with different participants for the trust game, and the (different rounds of) the SVO task, in order to eliminate reciprocity concerns.

[5] Our pre-registration form can be found on:
https://osf.io/792p4/?view_only=ca1b5bb7acdc4f5e9ec4b48bfdf509ce.

[6] The investment task was included in our design since an earlier version of our design used the standard trust game where the trustor's choice of how much to send can be affected by her risk preferences. Since this channel was eliminated in the final version of our trust game, we exclude the investment task from our analysis.

Die-rolling task Subjects are presented with the video of a randomly chosen die-roll on the screen and are asked to report the outcome for three rounds (Kocher et al., 2018). In each round they earn 75 points if they report a 5, and 0 points if they report any other number. The minimum amount a subject can earn from the die-rolling task is therefore 0 points, and the maximum is 225 points. We picked the die-rolling task since it is a simple task and it has been shown to correlate with cheating in several real-life situations (Cohn and Maréchal, 2018; Dai et al., 2018; Hanna and Wang, 2017; Potters and Stoop, 2016).

We measure the lying aversion of subjects as the ratio of the number of times they did not misreport a die-roll to the number of times they could have misreported to increase their earnings over the three rounds of the die-rolling task. Our measure of lying aversion is therefore defined as:

$$\text{lying aversion}_i = \frac{\#\text{of times subject } i \text{ could have gained from misreporting, but did not}}{\#\text{of times subject } i \text{ could have gained from misreporting}}$$

If the result of each die roll is a 5, there is no room for the subject to lie in order to increase their earnings. In this case, the denominator in our measure of lying aversion takes the value of 0, which makes the lying aversion not defined. This case occurred twice in our experiment, and we omit those observations in our analysis. We assume that subjects would only have an incentive to lie up, i.e., they would lie only when they can increase their earnings by doing so. Our lying aversion measure therefore takes values between 0 and 1, decreases with the number of times subjects misreport a die-roll, and hence, increases with the lying aversion of the subject. A subject with a lying aversion measure of 0 and a subject with a lying aversion measure of 1 are the least and the most honest extremes within our measure, respectively. We exclude the one subject who reported down, i.e., who observed a 5 but reported another number, from our analysis.

Partner choice and trust game We begin by outlining the order of events in our trust game in the no communication treatment. Afterwards we will describe how our treatments differ in detail. Within a session, we first randomly assign one third of the subjects to the role of trustor. The remaining subjects are assigned the role of trustee. We then split the trustees within the session into two groups, where one group includes the upper half and the other includes the lower half of the lying aversion distribution. Each trustor is then matched with two trustees, one randomly drawn from the first group, and the other randomly drawn from the second group. This is done with the purpose of eliminating uninformative observations by excluding trustee pairs with the same lying aversion measures as much as possible. We can however not always avoid trustee pairs with the same level of lying aversion due to limited session sizes. The results presented here exclude those cases. Since most of our subjects have lying aversion measures of either 0 or 1, the majority of the trustee pairs presented to trustors have a lying aversion difference of 1 (see Figure 1 in the main text).

In the trust game, the trustors are endowed with 50 points and have to pick one of the trustees they are matched with to send these 50 points to. Before picking their partner, they observe the die rolls and the reports of the two trustees from the die-rolling task. The trustors can therefore easily deduce which trustee is more lying averse. Upon sending, the 50 points are multiplied, either by 2 or by 4, where both multipliers are equally likely. The trustees, but not the trustors, observe the multiplier that is drawn. Both trustees choose how many points they want to send back to the trustor, if they are the chosen partner.

While trustees make their choices, we simultaneously elicit the trustors' beliefs on what they expect each trustee to return if the multiplier is 2, and if it is 4. Each trustor therefore submits four beliefs, two of which correspond to the true multiplier. We incentivize the trustors' beliefs for accuracy. For each trustee, trustors have a chance of winning 50 points, in addition to their earnings from the game. The probability of winning the prize is determined by how close their expectation is to how much the trustee returns, for the true multiplier. More specifically, if the trustor submits a belief of X points for a trustee and a multiplier level, and the trustee actually returns Y points, then the probability p of winning 50 points is:

$$p = 1 - \left(\frac{|X - Y|}{\text{Trustor endowment} * \text{Multiplier}} \right)^{1/2} = 1 - \left(\frac{|X - Y|}{50 * \text{Multiplier}} \right)^{1/2}$$

We then compare this probability to a random draw between 0 and 1, and if the probability p is higher than the random draw, the trustor wins the prize for that trustee; and if it is lower, the trustor does not win the prize. The trustors therefore has an incentive to submit their correct beliefs as it maximizes their chance of winning the prize.

After the trustees make their choices and the trustors submit their beliefs, the game ends. The trustees learn who is selected by the trustor, and the choice of that trustee is implemented: The trustor earns the amount sent back by the trustee she chose, and the chosen trustee earns the endowment she received times the multiplier minus how much she chose to send back. The non-chosen trustee earns the outcome of a random draw between 0 and 100, if the multiplier is 2; and between 0 and 200, if the multiplier is 4.

In the communication treatment, the trust game is played in exactly the same way, except that there is cheap-talk messaging. The trustees choose a message about the level of the multiplier to send to the trustor, together with the amount of points to send back. They choose between reporting that the multiplier was 2, or that it was 4. At the end of the trust game, the trustor observes the message sent by the trustee she chose, together with the amount sent back, and sends a reply. She chooses between saying that she does or does not believe the trustee. The reply for the trustor is included so that the trustor has *the last word*.

We take the share of points returned by the trustees and the share of points expected by the trustors as measures of actual and expected prosociality, respectively. Based on the existing literature and the mechanism outlined in the main text, we form the following hypotheses. Hypotheses 1-3 are about trustee behaviour. They test the relationship between prosociality, lying aversion and communication. Hypotheses 4-6 are about trustor beliefs. They test the relationship between expectations about partner's prosociality, partner's lying aversion and communication. And finally, Hypothesis 7 tests whether there is a preference for honest partners in the trust game.

- **Hypothesis 1 (H1):** There is a positive correlation between lying aversion and prosociality.
- **Hypothesis 2 (H2):** The existence of communication makes people behave more prosocially.
- **Hypothesis 3 (H3):** The impact of communication on prosocial tendencies is stronger for people with higher degrees of lying aversion.
- **Hypothesis 4 (H4):** People, in general, expect a positive correlation between lying aversion and prosociality.
- **Hypothesis 5 (H5):** People anticipate a positive impact of communication on prosocial tendencies.
- **Hypothesis 6 (H6):** People anticipate a stronger impact of communication on prosocial tendencies of the lying averse individuals.
- **Hypothesis 7 (H7):** The more lying averse people are chosen more often as a partner.

The impact of communication in H2-3 and H5-6 is anticipated in the cases where the multiplier is high. If the multiplier is 2, there is no (apparent) reason to misreport the multiplier. However, if the multiplier is 4, there is a tension between being prosocial, and being selfish and lying about it.

Analyses We have two types of statistical tests on trustee behaviour and trustor expectations. We run ordinary least squares (OLS) regressions, and we report Mann-Whitney tests results as a non-parametric check. We use robust standard errors in all of our regressions. Moreover, we cluster errors at the trustor level in regressions for trustor expectations since each trustor submits four separate beliefs. We run logistic (logit) regressions for studying the partner choice behaviour of trustors.

Social value orientation (SVO) measure We elicit subjects' SVO using the slider task (Crosetto et al., 2019; Murphy et al., 2011). Subjects make six resource allocation decisions. The alignment between payoffs for themselves and the other participant they are matched with is varied across decisions. The SVO angles calculated from the allocation decisions are used as a control variable of subjects' prosocial attitudes outside of the trust game.

3A.2 Results

3A.2.1 Treatment assignment

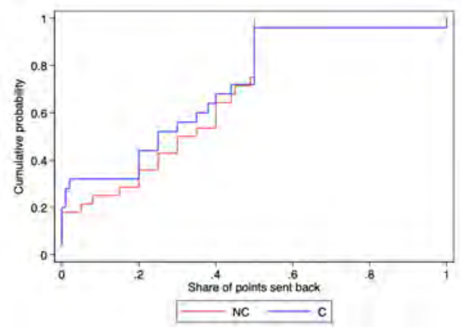
The distributions of subject attitudes across treatments, presented in Table 3A.1, suggest that our treatment assignment was random regarding subject characteristics.

Treatment	Lying aversion	SVO	Gender
C	0.535 (0.459)	16.492 (12.982)	0.522 -
NC	0.565 (0.450)	16.149 (13.245)	0.565 -

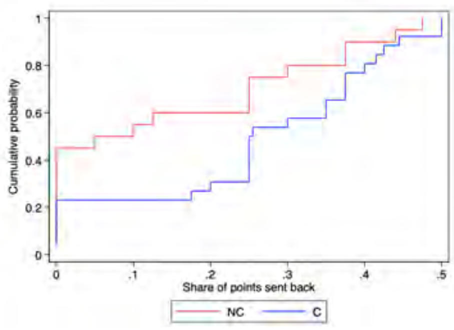
Table 3A.1: Summary statistics of subject characteristics across treatments: Mean values and standard deviations in parentheses. Mann-Whitney tests give a p-value of 0.6431 for the lying aversion comparison, and a p-value of 0.9354 for SVO angle comparison across treatments.

3A.2.2 Impact of communication

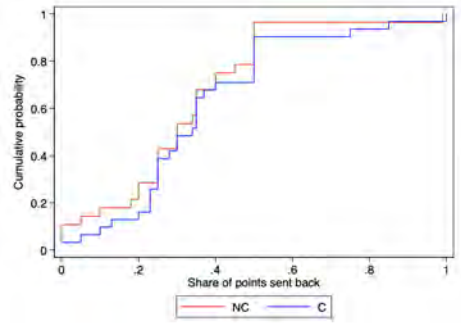
Distribution of trustee returns across treatments can be found in Figure 3A.1, and the distribution of trustor expectations across treatments can be found in Figure 3A.2.



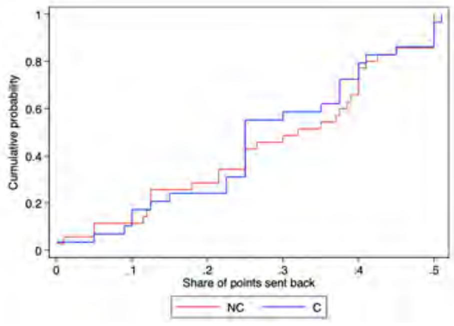
(a) Low multiplier & lying aversion = 0



(b) High multiplier & lying aversion = 0

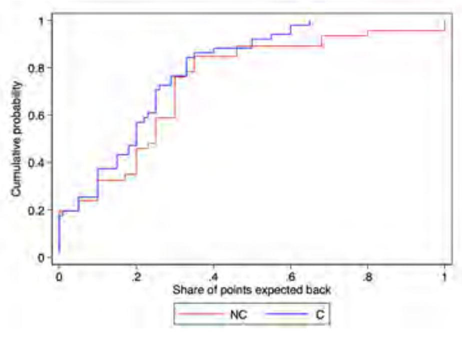


(c) Low multiplier & lying aversion = 1

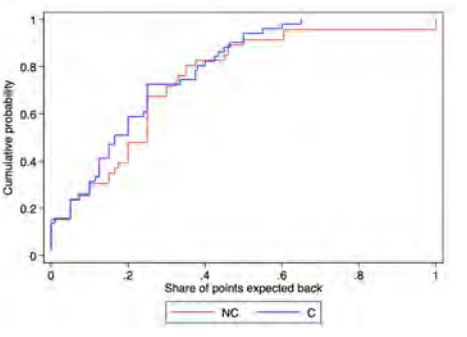


(d) High multiplier & lying aversion = 1

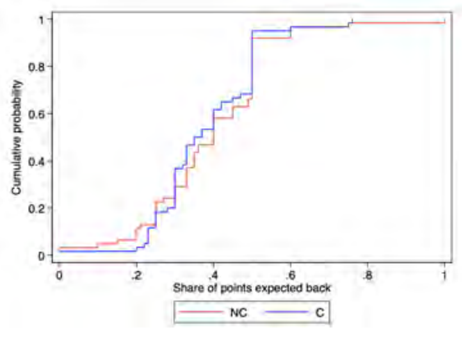
Figure 3A.1: Cumulative distribution functions of the share of points sent back by the trustees across treatments, for the cases with (a) low multiplier and trustees with a lying aversion measure of 0, (b) high multiplier and trustees with a lying aversion measure of 0, (c) low multiplier and trustees with a lying aversion measure of 1, and (d) high multiplier and trustees with a lying aversion measure of 1.



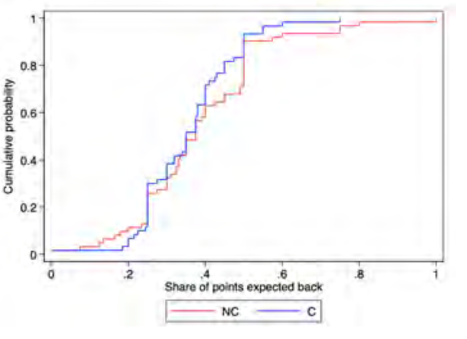
(a) Low multiplier & lying aversion of trustee = 0



(b) High multiplier & lying aversion of trustee = 0



(c) Low multiplier & lying aversion of trustee = 1



(d) High multiplier & lying aversion of trustee = 1

Figure 3A.2: Cumulative distribution functions of the share of points expected back by the trustors across treatments, for the cases with (a) low multiplier and trustees with a lying aversion measure of 0, (b) high multiplier and trustees with a lying aversion measure of 0, (c) low multiplier and trustees with a lying aversion measure of 1, and (d) high multiplier and trustees with a lying aversion measure of 1.

	(1) Misreport	(2) Misreport
Lying av.	-0.00587	-0.0347
	(0.992)	(0.950)
Constant	-0.636	-0.524
	(0.123)	(0.172)
N	55	68

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.2: Connection between misreporting the multiplier in the trust game and misreporting in the die-rolling task Logistic regression results on the probability that a trustee misreports the multiplier in the communication treatment when the multiplier is high, based on her lying aversion measure from the die-rolling task. Column I restricts data to trustees with a lying aversion of 0 or 1, Column II includes all trustees.

3A.2.3 Reports on the multiplier

Out of a total of 134 trustees the communication treatment, 102 trustees reported the multiplier truthfully, and 32 trustees misreported it. Out of 68 trustees who observed the high multiplier, 25 trustees misreported it. If we restrict observations to trustees with a lying aversion of 0 or 1, as done in the main text, we see that 19 out of 55 trustees misreported the multiplier. There does not seem to be a relation between misreporting the multiplier down in the trust game with misreporting the die-rolling outcomes in the die-rolling task (Table 3A.2). Out of 66 trustees who observed the low multiplier, 7 trustees misreported it upwards.

3A.3 Robustness checks on regression results

The tables below contain robustness checks we performed on our results. Tables 3A.1 and 3A.2 report results using alternative specifications for dividing trustee population as lying averse and not lying averse, for trustee returns and trustor expectations, respectively. Tables 3A.3 and 3A.4 report results including the full set of the control variables, for trustee returns and trustor expectations, respectively.

We measure the lying aversion of subjects with the die-rolling task. In order to eliminate the luck component in the die-rolling task, without it taking extremely long, we let subjects perform the task for three rounds. An individual's lying aversion measure from the die-rolling task might however still be blurred to some extent by their luck. For instance, consider a subject, who has not observed any 5s but reported truthfully in the first two rounds of the die-rolling task, but observes a 5 in the third round. In this case, she ends up with a positive payoff without any lying; however, she might have lied to earn some money if she had not observed a 5 in the last round. In this case, she would be classified as someone with a lying aversion of 1, whereas she could have instead been classified as someone with a lying aversion of 0. We therefore re-did our analyses (i) by including the number of 5s observed by a trustee as an independent variable, and (ii) by restricting the data to the subset of trustees who did not observe any 5s, in order to eliminate this type of noise. Tables 3A.5 and 3A.6 report results controlling for the number of 5s observed by trustees, for trustee returns and trustor expectations, respectively. Tables 3A.7 and 3A.8 report results including only the trustees who never observed a 5, for trustee returns and trustor expectations, respectively.

The dependent variables in our analysis, trustee returns and trustor expectations, are between 0 and 1. Tables 3A.9 and 3A.10 report results for the fractional response model specification, for trustee returns and trustor expectations, respectively; as a robustness check using a method specifically developed for bounded dependent variables. Finally, Tables 3A.11 and 3A.12 report results excluding outliers who sent/expected back more than 50%, for trustee returns and trustor expectations, respectively.

	(1) Ratio	(2) Ratio	(3) Ratio	(4) Ratio	(5) Ratio	(6) Ratio
C	0.0379 (0.313)		0.0389 (0.301)		0.0366 (0.331)	
Multiplier	-0.0601* (0.053)		-0.0538* (0.081)		-0.0635** (0.045)	
(C=1)#(M=1)	0.0291 (0.606)		0.0123 (0.810)		0.0299 (0.616)	
Lying av. continuous	0.0826*** (0.009)					
(C=1)#(M=1)#(LA)	-0.0544 (0.296)					
(C=0)#(M=1)#(1-LA)		-0.147*** (<0.001)				
Lying av. 1 vs. not 1			0.0601** (0.029)			
(C=1)#(M=1)#(LA=1)			-0.0404 (0.376)			
(C=0)#(M=1)#(LA=0)				-0.101*** (<0.001)		
Lying av. 0 vs. not 0					0.0707** (0.019)	
(C=1)#(M=1)#(LA=1)					-0.0444 (0.366)	
(C=0)#(M=1)#(LA=0)						-0.146*** (<0.001)
Constant	0.252*** (<0.001)	0.300*** (<0.001)	0.268*** (<0.001)	0.299*** (<0.001)	0.254*** (<0.001)	0.295*** (<0.001)
N	272	272	272	272	272	272
adj. R^2	0.045	0.045	0.034	0.032	0.040	0.038

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.1: OLS regression results on trustee returns with the alternative comparisons of trustees' lying aversion measures. Columns 1-2 include the continous measure of lying aversion, Columns 3-4 include the dummy variable measuring whether the trustee's lying aversion is 1 or whether it is lower than 1, and Columns 5-6 include the dummy variable measuring whether the trustee's lying aversion is 0 or whether it is higher than 0. This implies that Columns 3-4 separate individuals as those who never overreported and those who overreported at least once, and Columns 5-6 separate individuals as those who always overreported and those who reported truthfully at least once. In Columns 1-2, LA is the actual level of the lying aversion measure. In Columns 3-6, LA=1 and LA=0 indicate the values of the indicator variable for the corresponding lying aversion comparison.

	(1)	(2)	(3)	(4)	(5)	(6)
	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio
C	-0.0265	-0.0253	-0.0298	-0.0287	-0.0272	-0.0261
	(0.287)	(0.319)	(0.226)	(0.253)	(0.278)	(0.309)
Multiplier	-0.00679	-0.00567	-0.00679	-0.00567	-0.00679	-0.00567
	(0.270)	(0.388)	(0.270)	(0.388)	(0.270)	(0.388)
(C=1)#(M=1)		0.000438		0.000690		-0.00333
		(0.983)		(0.969)		(0.880)
Lying av. continuous	0.150***	0.152***				
	(<0.001)	(<0.001)				
(C=1)#(M=1)#(LA)		-0.00500				
		(0.866)				
Lying av. 1 vs. not 1			0.125***	0.127***		
			(<0.001)	(<0.001)		
(C=1)#(M=1)#(LA=1)				-0.00654		
				(0.798)		
Lying av. 0 vs. not 0					0.121***	0.120***
					(<0.001)	(<0.001)
(C=1)#(M=1)#(LA=1)						0.00176
						(0.953)
Constant	0.244***	0.242***	0.272***	0.270***	0.250***	0.250***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N	536	536	536	536	536	536
adj. R^2	0.141	0.138	0.119	0.115	0.103	0.099

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.2: OLS regression results on trustor expectations with the alternative comparisons of trustees' lying aversion measures. Columns 1-2 include the continous measure of lying aversion, Columns 3-4 include the dummy variable measuring whether the trustee's lying aversion is 1 or whether it is lower than 1, and Columns 5-6 include the dummy variable measuring whether the trustee's lying aversion is 0 or whether it is higher than 0. This implies that Columns 3-4 separate individuals as those who never overreported and those who overreported at least once, and Columns 5-6 separate individuals as those who always overreported and those who reported truthfully at least once. In Columns 1-2, LA is the actual level of the lying aversion measure. In Columns 3-6, LA=1 indicates the value of the indicator variable for the corresponding lying aversion comparison.

	(1) Ratio	(2) Ratio	(3) Ratio
Lying av.	0.0288	0.0405	
	(0.281)	(0.226)	
C	0.0254	0.0253	
	(0.317)	(0.532)	
Multiplier	-0.0588**	-0.0600*	
	(0.017)	(0.079)	
SVO angle	0.00398***	0.00397***	0.00413***
	(<0.001)	(<0.001)	(<0.001)
Female	-0.0514**	-0.0528**	-0.0501*
	(0.050)	(0.049)	(0.053)
Age	0.00115	0.00123	0.00107
	(0.723)	(0.707)	(0.739)
Experience1	-0.0187*	-0.0179	-0.0178*
	(0.087)	(0.111)	(0.099)
Experience2	0.00966	0.00812	0.00950
	(0.520)	(0.597)	(0.523)
Europe	-0.0597**	-0.0603**	-0.0575**
	(0.033)	(0.035)	(0.030)
Econ	-0.0192	-0.0183	-0.0179
	(0.498)	(0.523)	(0.511)
(C=1)#(M=1)		0.0233	
		(0.709)	
(C=1)#(M=1)#(LA=1)		-0.0426	
		(0.412)	
(C=1)#(M=1)#(LA=0)			-0.125***
			(0.001)
Constant	0.293***	0.285***	0.299***
	(0.001)	(0.002)	(<0.001)
N	222	222	222
adj. R^2	0.134	0.128	0.146

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.3: OLS regression results on trustee returns including control variables. SVO angle is the SVO angle of the subjects based on their choices in the SVO task. Female is a dummy variable that takes the value of 1 for females. Age is the age of the subject. Experience1 and Experience2 are subjects' experience level in experiments in the CREED lab and outside, respectively. Europe is a dummy variable indicating European nationality. Econ is a dummy variable indicating whether the subject majors in a department in the Economics and Business Faculty.

	(1) Ratio	(2) Ratio
Lying av.	0.143***	0.144***
	(<0.001)	(<0.001)
C	-0.0255	-0.0238
	(0.343)	(0.394)
Multiplier	-0.00842	-0.00676
	(0.206)	(0.378)
Lying av. trustor	0.0120	0.0121
	(0.744)	(0.744)
SVO angle	0.00124	0.00124
	(0.241)	(0.241)
Female	0.00415	0.00419
	(0.883)	(0.882)
Age	0.00341**	0.00341**
	(0.038)	(0.038)
Experience1	-0.0234*	-0.0234*
	(0.063)	(0.065)
Experience2	-0.0308	-0.0308
	(0.159)	(0.161)
Europe	-0.0137	-0.0137
	(0.697)	(0.698)
Econ	-0.0191	-0.0190
	(0.643)	(0.645)
(C=1)#(M=1)		-0.00144
		(0.945)
(C=1)#(M=1)#(LA=1)		-0.00341
		(0.909)
Constant	0.206**	0.204**
	(0.024)	(0.028)
N	438	438
adj. R^2	0.175	0.171

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.4: OLS regression results on trustor expectations including control variables. SVO angle is the SVO angle of the subjects based on their choices in the SVO task. Female is a dummy variable that takes the value of 1 for females. Age is the age of the subject. Experience1 and Experience2 are subjects' experience level in experiments in the CREED lab and outside, respectively. Europe is a dummy variable indicating European nationality. Econ is a dummy variable indicating whether the subject majors in a department at the Economics and Business Faculty.

	(1) Ratio	(2) Ratio	(3) Ratio
Lying av.	0.0639**	0.0779**	
	(0.014)	(0.015)	
C	0.0278	0.0168	
	(0.279)	(0.684)	
Multiplier	-0.0619**	-0.0741**	
	(0.016)	(0.038)	
No. of 5s	0.00694	0.00770	0.00824
	(0.708)	(0.678)	(0.656)
(C=1)#(M=1)		0.0503	
		(0.424)	
(C=1)#(M=1)#(LA=1)		-0.0523	
		(0.325)	
(C=1)#(M=1)#(LA=0)			-0.149***
			(<0.001)
Constant	0.263***	0.260***	0.294***
	(<0.001)	(<0.001)	(<0.001)
N	222	222	222
adj. R^2	0.038	0.033	0.041

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.5: OLS regression results on trustee returns controlling for the number of 5s observed by the trustee.

	(1)	(2)
	Ratio	Ratio
Lying av.	0.144***	0.145***
	(<0.001)	(<0.001)
C	-0.0306	-0.0289
	(0.267)	(0.313)
Multiplier	-0.00842	-0.00676
	(0.202)	(0.375)
No. of 5s	0.00543	0.00546
	(0.756)	(0.756)
(C=1)#(M=1)		-0.000837
		(0.968)
(C=1)#(M=1)#(LA=1)		-0.00453
		(0.880)
Constant	0.250***	0.249***
	(<0.001)	(<0.001)
N	438	438
adj. R^2	0.144	0.140

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.6: OLS regression results on trustor expectations controlling for the number of 5s observed by the trustee.

	(1) Ratio	(2) Ratio	(3) Ratio
Lying av.	0.0730**	0.0888**	
	(0.043)	(0.041)	
C	0.0396	0.0000379	
	(0.265)	(0.999)	
Multiplier	-0.0579*	-0.101**	
	(0.095)	(0.024)	
(C=1)#(M=1)		0.111	
		(0.186)	
(C=1)#(M=1)#(LA=1)		-0.0524	
		(0.463)	
(C=1)#(M=1)#(LA=0)			-0.187***
			(<0.001)
Constant	0.251***	0.263***	0.298***
	(<0.001)	(<0.001)	(<0.001)
N	120	120	120
adj. R^2	0.037	0.035	0.068

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.7: OLS regression results on trustee returns, restricted to the group of trustees who never observed a 5.

	(1) Ratio	(2) Ratio
Lying av.	0.191***	0.189***
	(<0.001)	(<0.001)
C	-0.0770**	-0.0710*
	(0.041)	(0.067)
Multiplier	-0.0134*	-0.00725
	(0.085)	(0.444)
(C=1)#(M=1)		-0.0171
		(0.503)
(C=1)#(M=1)#(LA=1)		0.00946
		(0.817)
Constant	0.252***	0.250***
	(<0.001)	(<0.001)
N	236	236
adj. R^2	0.253	0.247

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.8: OLS regression results on trustor expectations, restricted to the group of trustees who never observed a 5.

	(1)	(2)	(3)
	Ratio	Ratio	Ratio
Lying av.	0.194**	0.237**	
	(0.013)	(0.014)	
C	0.0844	0.0468	
	(0.260)	(0.684)	
Multiplier	-0.183**	-0.227**	
	(0.014)	(0.033)	
(C=1)#(M=1)		0.165	
		(0.384)	
(C=1)#(M=1)#(LA=1)		-0.154	
		(0.335)	
(C=0)#(M=1)#(LA=0)			-0.510***
			(0.002)
Constant	-0.632***	-0.636***	-0.529***
	(<0.001)	(<0.001)	(<0.001)
N	222	222	222

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.9: Fractional response regression results on trustee returns.

	(1)	(2)
	Ratio	Ratio
Lying av.	0.419***	0.418***
	(<0.001)	(<0.001)
C	-0.0885*	-0.0839
	(0.071)	(0.229)
Multiplier	-0.0238	-0.0192
	(0.627)	(0.802)
(C=1)#(M=1)		-0.0118
		(0.927)
(C=1)#(M=1)#(LA=1)		0.00463
		(0.968)
Constant	-0.673***	-0.675***
	(<0.001)	(<0.001)
N	438	438

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.10: Fractional response regression results on trustor expectations.

	(1) Ratio	(2) Ratio	(3) Ratio
Lying av.	0.0494**	0.0623**	
	(0.035)	(0.027)	
C	0.0104	-0.0133	
	(0.648)	(0.696)	
Multiplier	-0.0351	-0.0594*	
	(0.124)	(0.073)	
(C=1)#(M=1)		0.0701	
		(0.217)	
(C=1)#(M=1)#(LA=1)		-0.0428	
		(0.397)	
(C=1)#(M=1)#(LA=0)			-0.132***
			(0.001)
Constant	0.255***	0.260***	0.282***
	(<0.001)	(<0.001)	(<0.001)
N	216	216	216
adj. R^2	0.017	0.016	0.047

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.11: OLS regression results on trustee returns, excluding the outliers who sent back more than 50% of the available amount.

	(1) Ratio	(2) Ratio
Lying av.	0.164***	0.170***
	(<0.001)	(<0.001)
C	-0.00710	-0.00635
	(0.703)	(0.748)
Multiplier	-0.00610	-0.00546
	(0.380)	(0.520)
(C=1)#(M=1)		0.0110
		(0.616)
(C=1)#(M=1)#(LA=1)		-0.0225
		(0.414)
Constant	0.195***	0.192***
	(<0.001)	(<0.001)
N	404	404
adj. R^2	0.285	0.283

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A.12: OLS regression results on trustor expectations, excluding the outliers who expected back more than 50% of the available amount.

3A.4 Instructions

Our instructions can be found below. The subjects received the instructions on their computer screens. In the trust game, each subject got instructions from the perspective of their role in the trust game on their screen. Subjects were also handed out printed copies of the general instructions for the die-rolling task and the trust game. Unless stated otherwise, the instructions were the same for both treatments. The instructions for the die-rolling task are based on Gross et al. (2018), and the instructions for the SVO task are based on Böhm et al. (2018).

3A.4.1 Die-rolling instructions

Part 1 of the experiment consists of 3 rounds of a die rolling task.

- In each round, you will see a random roll of a die on the computer screen.
- Your task is to report the outcome of the die roll that you saw on the screen.
- Your payoff will be determined by the result that you report.
- You will be paid for all of your reports. Specifically, for each round, ...
  - ... you will earn 75 points if you report a 5.
  - ... you will earn 0 points otherwise.

Figure 3A.1: A screenshot of the die-roll task with the videos created by Kocher et al. (2018).

3A.4.2 Trust game instructions

In Part 2 of the experiment, you will interact with two other participants. In this interaction, there are three different roles: *Participant A, Participant B, and Participant C*

The order of events is as follows:

1. Participant A is given 50 points. Participant A observes the die rolls and the reports of Participant B and Participant C from Part 1 of the experiment. Participant A then chooses to send the 50 points either to Participant B or to Participant C. Until the end of Part 2, Participant B and Participant C will not learn who was picked by Participant A.

2. Upon sending, the 50 points are multiplied by 2 or by 4. The multiplier is chosen randomly and both multipliers are equally likely. Participant B and Participant C both observe the multiplier. Participant A does not observe the multiplier.

3. Participant B and Participant C both choose how many points they want to send back to Participant A, in case they are chosen. Participant B and Participant C can choose to send back any amount between 0 points and the full amount available, which is 100 points if the multiplier is 2, and 200 points if the multiplier is 4.

4. **(Only in the communication treatment)** Together with the amount to be sent back, Participant B and Participant C will also choose a message about the multiplier. They can choose to say to Participant A that the multiplier was 2 or that it was 4.

5. The choices of the participant that Participant A selected are implemented. Participant A observes the amount sent back by the participant that s/he chose. **(Only in the communication treatment)** Participant A also observes the message sent by the participant that s/he chose.

6. **(Only in the communication treatment)** Participant A then sends a reply to the chosen participant. S/he can either reply that s/he believes, or that s/he does not believe the message sent by this participant.

7. Participant B and Participant C learn who was chosen by Participant A. **(Only in the communication treatment)** The participant that was chosen by Participant A then observes the reply that Participant A sent to her/him.

The payoffs in Part 2 are determined as follows:

- Participant A earns the amount sent back by the participant that s/he chose.
- The participant that Participant A chose earns the amount s/he received minus the amount s/he sent back to Participant A.
- The participant that was not chosen earns an amount that is determined by a random draw. The range for this random draw depends on the multiplier. Specifically, ...

– ... if the multiplier is 2, the non-chosen participant earns an amount ranging between 0 and 100 points, with increments of 1 point. Any of these possible payoffs is drawn with equal probability.

– ... if the multiplier is 4, the non-chosen participant earns an amount ranging between 0 and 200 points, with increments of 1 point. Any of these possible payoffs is drawn with equal probability.

Belief elicitation instructions In the next page, you will be asked to predict how many points Participants B and C will choose to send back to you. Both for Participant B and for Participant C, we will ask you to predict what that participant will send back in case that the multiplier is 2, and in case that the multiplier is 4. So, you will submit four predictions in total.

Based on your predictions about each participant, you will have a chance of winning a prize of 50 points. The closer your prediction is to the actual choice of the participant, the higher your chance of winning the prize will be.

More specifically, given the actual value of the multiplier, we will compare the amounts that Participant B and Participant C (would have) sent back with the amounts that you predicted they would send back for that value of the multiplier. We then use a formula to determine your chance of winning the prize. The formula is designed such that your chance of winning the prize becomes higher, the closer your prediction is to the actual choice of the participant.

Note that the formula ensures that you maximize your chance of winning money by stating your true beliefs regarding the choices of Participant B and Participant C for each level of the multiplier. It is not important that you understand the formula in detail. What matters is for you to know that you maximize your chances of winning by truthfully reporting your best estimate regarding the other participants' choices. If you under- or overstate your true beliefs, you will reduce your chances of winning the prize.

However, if you want to take a closer look at the formula, we have prepared some more detailed information for you. Below you can choose to proceed straight to the prediction task, or to first review the additional information on the formula.

Do you want to review the details of the formula?

[IF YES]

How a given answer translates to your chance of winning the prize of 50 points is based on a formula. The formula is designed to make sure that you maximize your chance of winning if you truthfully report your best estimates of the amounts you expect Participant B and Participant C to send back for each level of the multiplier.

Suppose that the multiplier is 2, Participant B chooses to send you R_1 points and Participant C chooses to send you S_1 points. The variables r_1 and s_1 are your reports for this case - your reports on how much you think Participants B and C will choose to send back to you in case the multiplier is 2. The winning probabilities $p_{1,B}$ for your prediction on Participant B and $p_{1,C}$ for your prediction on Participant C are then given by:

$$p_{1,B} = 1 - \left(\frac{|r_1 - R_1|}{100}\right)^{1/2} \quad \text{and} \quad p_{1,C} = 1 - \left(\frac{|s_1 - S_1|}{100}\right)^{1/2}$$

Similarly, for the case where the multiplier is 4, the winning probabilities $p_{2,B}$ for your prediction on Participant B and $p_{2,C}$ for your prediction on Participant C are then given by:

$$p_{2,B} = 1 - \left(\frac{|r_2 - R_2|}{200}\right)^{1/2} \quad \text{and} \quad p_{2,C} = 1 - \left(\frac{|s_2 - S_2|}{200}\right)^{1/2}$$

All of the p-functions attain the value of 1 if, and only if, your guess is equal to the actual amount sent by (respectively) Participant B or Participant C. Moreover, using some calculus, one can show that each of the p-functions is strictly increasing below the actual amount sent by (respectively) Participant B or Participant C, and it is strictly decreasing above that amount. So, while you do not know the actual amount each participant decided to send back for a given amount of the multiplier, the best thing you can do to make your probability of winning as high as possible, is to submit the average amount that you expect a participant like Participant B, or like Participant C, to return.

Below you can find a simple example. Suppose that the multiplier is 2, and that you believe that in this case, Participant B would choose to send you 25 points with probability 0.1, 33 points with probability 0.8, and 41 points with probability 0.1. In the table below, you can find the probability of winning the prize of 50 points for different beliefs you can choose to report about Participant B's choice for this case.

For example, if you report that you expect Participant B to send back 25 points, the probability that you win the prize is 73.37%. As you can see, you maximize your chance of winning by reporting the average amount that you expect Participant B to return.

	Report #1	Report #2	Report #3	Report #4	Report #5
Hypothetical report r_1	12	25	33	41	56
Expected winning probability $p_{1,B}$	54.35%	73.37%	94.34%	73.37%	52.19%

3A.4.3 Social value orientation task instructions

- In Part 4, you will be making a series of 6 decisions about allocating resources between yourself and a randomly selected participant.
- The participant you will be matched with will be a different person than the participants you were matched with in Part 2.
- In every decision, a slider bar will represent the allocations that are available to you. You select your most preferred allocation by marking the corresponding position on the slider bar.
- At the end of the experiment, we will randomly pick one of the decisions from Part 4 for payment.
- You will be paid twice, once being in the *sender role* and once being in the *receiver role*:
  - In the *sender role*, you get the points you decided to keep on the selected decision according to your own allocation decision. A randomly selected participant will be in the receiving role for your allocation decision.
  - In the *receiver role*, you will receive points based on the allocation decision of a second randomly chosen participant for that decision. That participant will be different from the person that received points based on your allocation decision and all other participants you were previously matched with in the experiment.

3A.4.4 Survey items

Please answer the following questions:

- What is your age?
- What is your gender?
- What are you studying?
- Where are you from?

- How many experiments have you participated at CREED (excluding this experiment)? Please select the right category below.
- How many experiments have you participated at other institutes than CREED? Please select the right category below.
- Reminder: In Part 1 of the experiment, you did a die rolling task for 3 rounds. In each round, you observed a random roll of a die on the computer screen and reported the outcome of the die roll. You can see the rolls you observed and the outcomes you reported in the table below.

  [REMINDER ON PART I BEHAVIOUR]

  How did you make your choices in Part 1? Could you describe the method(s) you followed?
- Reminder: In Part 2 of the experiment, you interacted with two other participants. In this interaction, Participant A chose either Participant B or Participant C, and sent 50 points to that participant. These points were then multiplied by 2, or by 4. Participants B and C observed the multiplier and chose an amount to send back to Participant A.

  (The communication treatment) Together with the amount to be sent back, Participants B and C also chose a message about the multiplier. The choices of the participant that was chosen by Participant A were implemented. Participant A then observed the amount and the message sent by the participant that s/he chose.

  (The no communication treatment) The choice of the participant that was chosen by Participant A was implemented. Participant A then observed the amount sent by the participant that s/he chose.

  [REMINDER ON PART II ROLE AND BEHAVIOUR]

  How did you make your choices in Part 2? Could you describe the method(s) you followed?