Are we more honest than others think we are?

Claire Mouminoux, Jean-Louis Rullière

Are we more honest than others think we are?

Claire Mouminoux, Jean-Louis Rullière

June 30, 2021

Abstract

We here examine the relationship between own honesty and beliefs about others' honesty under more- or less- favorable conditions. In a laboratory experiment, unfavorable conditions lead both participants to become more dishonest and reduce their beliefs in others' honesty. We find that participants are less honest than others think they are. In particular, participants underestimate the degree of others' dishonesty in unfavorable environments. Our results in addition show that dishonest participants believe themselves that others are.

JEL classification: C91, D01, D84.

Keywords— honesty; beliefs; experiments, behavioral economics

1 INTRODUCTION

Honesty and beliefs about others' honesty are central in many economic and social interactions. While Laws are justified on the basis of the benefits they yield for society, policy-makers focus on the reasons for which some people violate the Law. The answer most often put forward is that given by Becker (1968). In his seminal article on crime and punishment, he assumed that honest or dishonest behavior results from the comparison of the expected pecuniary costs to the associated benefits. This work has paved the way for many contributions focusing on the role of monetary incentives in honesty, but nonetheless without reaching a consensus. Some work has uncovered a negative relationship between monetary stakes and dishonesty (Balasubramanian et al. 2017, Cohn et al. 2019), some a positive relationship (Kajackaite & Gneezy 2017), and a last group no relationship between the two (Mazar et al. 2008, Fischbacher & Follmi-Heusi 2013, Andersen et al. 2018). Gibson et al. (2013), Gneezy et al. (2013) and Gneezy et al. (2018) all highlight the comparison of honest and dishonest earnings.

In their meta-analysis, Abeler et al. (2019) conclude that the consequences of higher potential payoffs are limited. As such, the decision to cheat likely depends not only on the pecuniary consequences but also on many other characteristics. For example, Kajackaite & Gneezy (2017) argue that when lying is an explicit rule of the game, subjects place more importance on the possibility of being caught out as a liar. A greater monetary incentive is therefore required to compensate for the cost of detection. In line with this idea, Yaniv & Siniver (2016) show that individuals in an unmonitored environment cheat considerably. This previous work seems to indicate that individuals' honesty is mainly driven by the consequences of their actions on their social ethics. In addition, both Houser et al. (2012) and Galeotti et al. (2017) highlight the importance of the perception of a given environment on honesty. Subjects are less honest in situations that are perceived to be unfair, with unfairness dampening their social-ethics considerations in their honesty decisions.

Above economic calculations and social-ethics considerations, research on trust and trustworthiness underlines important effects of social interaction on the willingness to trust others. Berg et al. (1995) underline the importance of information on trust reciprocity, showing that participants who are given information about others or who have previously interacted with them are more likely to trust and to be trusted in turn. In the same vein, in Glaeser et al. (1996, 2000) individuals who share social characteristics and have frequent social interactions are more likely to trust each other. Ermisch et al. (2009) analyze how population characteristics affect trust and trustworthiness by showing, for example, that individuals in "comfortable" situations are more likely to be trusting.

As previsouly demonstrated, a significant amount of research has been devoted to the analysis of honesty and individuals' motivations to lie, cheat or trust another party. But there is another indirect cost that is generally ignored: the fact that one person can violate the Law (whether she does so or not) can reduce beliefs in others' honesty, and lead others to become more dishonest in turn. In addition, underestimating others' honesty can lead to wasteful and costly checking, or missing out on valuable services. Why hire ticket inspectors if everyone travels with a valid ticket? On the contrary, overestimating others' honesty can lead to inefficient exchanges based on dishonest information, such as an insurance broker recommending the most-profitable rather than the most-appropriate contract. What is the relationship between own honesty and beliefs in others' honesty, and is one person's honesty accurately predicted by others?

To the best of our knowledge, little work has addressed the relationship between own honesty and beliefs in others' honesty. Hugh-Jones (2016) shows that at the aggregate level, beliefs in others' honesty do not match to observed honesty. He finds that the match between beliefs and honesty is best when considering participants in the same country, suggesting that beliefs may be driven by self-projection. However, this result can also be explained by the difference in subjects' knowledge about other countries that may influence their expectations. Maux et al. (2021) also find that beliefs and observed honesty are different when participants do not have information about others' previous actions. In addition, and similarly to Rauhut (2013), providing information about others' dishonesty leads participants who underestimated others' dishonesty to lie more, and those who overestimated it to lie less. This result suggests an important relationship between beliefs and

honesty behavior, but does not focus on the sources of any spread between the two.

Our experiment aims to measure three key characteristics: honesty, beliefs in others' honesty, and the effect of a more- or less-fortunate initial situation on the two. Our approach differs from that in previous works, as it produces data on honesty and beliefs in others' honesty at the individual level where the decision-making environment (a more-or less-favorable environment) is also observable at the individual level too. Participants are given a wallet containing five Euros. The rule of the game consists of a random draw indicating how many Euros they are allowed to take from the wallet. However, they can take as many as they wish up to the maximum of five Euros, and play anonymously without any monitoring. The draw can be more- or less-fortunate: an individual receiving the most-favorable draw is instructed to take all five Euros, and thus does not need to cheat. At the opposite end of the spectrum, the least-favorable draw instructs the participant to leave all of the money in the wallet. This produces experimental conditions in which every participant can take any amount of money from the wallet, but with different degrees of cheating for a given amount taken. Similar to Cohn et al. (2015), Fischbacher & Follmi-Heusi (2013), Pruckner & Sausgruber (2013), Nagin & Pogarsky (2003) and Mazar et al. (2008), we interpret honesty as the degree of compliance with the instructions.

We investigate the relationship between own honesty and own beliefs in others' honesty while accounting for whether the draw is more- or less-fortunate. As in Gibson et al. (2013), Gneezy et al. (2013) and Gneezy et al. (2018), participants with less-favorable draws are more likely to be dishonest. Lower earnings from honesty produce greater dishonesty and beliefs in others dishonesty at the extensive margin (the probability of being dishonest and believing that others are dishonest). On the contrary, less-favorable conditions do not affect the intensive margin of dishonesty (the extent of the dishonesty) but do reduce beliefs in the degree of others' dishonesty. Overall, we find that participants are less honest than others think they are.

At the extensive margin, there is no significant different between the percentage of participants who do not follow the game's rules (21.2%) and the percentage who expect others to cheat (21.9%). However, at the intensive margin participants significantly underestimate the degree of dishonesty. More precisely, the size of the spread is underestimated for the unfortunate draw and overestimated otherwise. Overall, underestimation predominates. Participants believe that others who are dishonest because they received an unfavorable draw will be less dishonest than they actually are at the intensive margin. We do, however find that a more- or less-favorable draw only affects the decision to be dishonest, and not the extent of the dishonesty.

We in addition find similar effects on participants' beliefs in their own (hypothetical) honesty. We include a self-assessment honesty survey, and uncover hypothetical biases in ex-ante self-assessments, so that participants underestimate their future dishonesty: honest participants overestimate their dishonesty under unfavorable draws, while dishonest participants underestimate their dishonesty under favorable draws. In addition, participants who said they would be dishonest are even more dishonest at the intensive margin than they declared.

Last, we find a strong positive correlation between individuals' own honesty and their beliefs in others' honesty. While there is likely two-way causality, our results suggest that participants are dishonest because they believe that others are dishonest. Robert & Arnab (2012) show that dishonesty is contagious when information about others' dishonesty is provided to participants. We here find that dishonesty is indirectly contagious through participants' beliefs, even under anonymity and without social interactions. Controlling for individual characteristics, we find that men are more likely to be dishonest (similarly to Grolleau et al. (2016)). However, there is no significant difference between men's and women's beliefs in others' honesty, so that men do not pass their own greater dishonesty on to their beliefs about others' behavior.

The remainder of this paper is organized as follows. In Section 2, we present our experimental design, and Section 3 briefly describes the sample. The results appear in Section 4. Last, Section 5 discusses and concludes.

2 EXPERIMENTAL DESIGN

We set up an original experiment to elicit participants' honesty and beliefs about others' honesty according to free compliance with an objective rule. The consequences of this rule differ from one participant to another depending on how favorable the draw is. There is no monitoring of compliance with the rule, and choices are anonymous. We have two different treatments to control for task order, framing effects and hypothetical biases.

2.1 An objective rule

The objective rule given to participants is relatively simple. There is a wallet containing 10 coins of 50 Euro Cents and a small piece of card showing the result of 10 independent draws. The 10 draws with replacement are made from a bag containing a white ball and a black ball.

The participant is instructed to take from the wallet the number of coins corresponding to the number of black balls displayed on the card.

The consequences of this rule therefore differ widely for participants receiving a draw from 10 black balls to 10 white balls. The former is instructed to take all 5 Euros from the wallet while the latter should, according to the rule, leave all of the coins inside.

2.2 Definition of honesty and beliefs in others' honesty

The second aim of the experiment is to determine the individual characteristics that lie behind own honesty and beliefs in others' honesty. To do so, we define:

- Own honesty = the difference between the amount taken from the wallet and the amount indicated by the rule;
- Beliefs in Others' Honesty = the difference between the amount the individual expects others to take from the wallet and the amount indicated by the rule.

2.3 Task implementation via two different treatments

We measure own honesty and beliefs in others' honesty in two treatments reflecting different ordering of the two tasks. These treatments are carried out in a within-group design.

- ***Behave+Believe* treatment**. In this treatment, we first elicit participants' honesty and then participants' beliefs in others' honesty.

  Participants take their places in the experimental laboratory and find a wallet on their table. On the top of this wallet, there is a three-letter code allowing them to enter a web interface to receive instructions and continue with the experiment after the honesty task. We first ask them to enter the three-letter code from the wallet without touching anything on the table (we also ask them to leave the code on the wallet, allowing us to match honesty to other participant characteristics), and to wait for the instructions.

  Three step-by-step instructions appear on their screens, and are read aloud at the same time by the experimenter:

> Instruction 1:
> There is a wallet and a padded envelope (empty) on the table of each participant. Please wait for our signal to reveal its contents. The wallet contains 10 50-cent coins (5 Euros in total) and a small piece of card showing the result of 10 independent draws between a white and a black ball.
> In this part, we ask you to apply the following rule:
>
> – For each black ball, you can take 50 cents from the wallet and put these 50 cents in the padded envelope.
> – The Euros left in the wallet correspond to 50 cents times the number of white balls.
>
> The experimenter and the other participants cannot observe you. You are not monitored and all of the wallets are put in the same bag indiscriminately at the end of this part of the experiment. For this part, your earnings correspond to the amount that you put in the padded envelope.
>
> Instruction 2:
> Please indicate below, in front of each piece of card:
>
> **How many Euros would you have taken if you had received this draw?**
>
> Instruction 3:
> During a previous experimental session, we distributed to participants a wallet together with a padded envelope (empty) similar to those that you have on your table. The wallet contained 10 50-cent coins (5 Euros in total) and a small piece of card showing the result of 10 independent draws between a white and a black ball.
>
> We asked participants to apply the following rule:
>
> – For each black ball, they should take 50 cents from the wallet and put these 50 cents in the

padded envelope.

– The Euros left in the wallet correspond to 50 cents times the number of white balls.

The experimenter and the other participants were not able to observe the participants' actions. They were not monitored and all of the wallets were put in the same bag indiscriminately at the end of this part of the experiment.
In this part, the different draws that were given to the participants will appear on your screen. You have to indicate for each of these:

How many Euros do you think the participant who received this draw took from the wallet?

To determine your earnings for this part, we will randomly select one of these draws and you will receive: 5 Euros −|your estimation error|.

- ***Believe+Behave* treatment**: In this treatment, we elicit participants' honesty and participants' beliefs in others' honesty. These treatments are performed using a within-group design. The procedure here is similar to that in the *Behave+Believe* treatment, except that *Instruction 1* and *Instruction 3* are inverted.

Paper versions of the instructions were distributed to participants at the same time, including examples and general information about the experiment. These instructions were in French, and we provide an English translation in the supplementary materials (see Section 3 (Procedure) for the supplementary-materials link).

3 PROCEDURE

The experiment was carried out using an original interface developed with HTML and JavaScript, with the back end with Java and PostgreSQL as the database. For full transparency, we have made available the experimental data and the R code used in this paper as supplementary material available.

The subjects were students from a French University at the Bachelor level. The sessions were carried out between June and December 2019, with a total of 165 subjects in the experiment. The experiment is without any context and that (dis)honesty is never mentioned, the authors and the experimentalist helpers have no connection with the participating students in order to avoid any demand effect. Table 1 summarizes the sample by the two treatments. There were 8 sessions, four for each of the two treatments (*Believe+Behave* and *Behave+Believe*).

Around 21 subjects took part in each session. The honesty and beliefs in others' honesty elicitation tasks were part of a larger-scale experiment. These two tasks were always played first, and the instructions were provided at the beginning of each part to avoid any framing effects. Before leaving the room to receive their payment privately, we asked participants to answer some general questions about their age, gender and education. The sessions lasted about 60 minutes, and the

honesty and beliefs in others' honesty elication tasks took on average 20 minutes. The total average payoff was around 10€, including a show-up fee of 3€. The total payoffs of the honesty elicitation and beliefs in others' honesty elicitation tasks were around 7€ (respectively around 3€ and 4€).

Table 1: Sample description by treatment

Treatment	N	Age		Gender	
		Mean	*SD*	*Female*	*Male*
Believe+Behave	88	18.4	0.77	54.5%	45.5%
Behave+Believe	77	19.2	1.76	66.2%	33.8%
Total	165	18.8	1.37	40%	60%

4 RESULTS

4.1 Definitions

For simplicity, we define the following metrics:

- The Allowed Amount according to the rule (**AA**): This is the amount (in Euros) that participants can take following the rule. It takes on values of 0 to 5, by units of 0.5.
- The Declared Withdrawn Amount (**DWA**): This is the answer to the question *How many Euros would you have taken if you had received this draw?* As a reminder, this question was asked between the two tasks.
- The Expected Withdrawn Amount (**EWA**): This is the answer to the question *How many Euros do you think the participant who received this draw took from the wallet?* It takes on values of 0 to 5, by units of 0.5.
- The Observed Withdrawn Amount (**OWA**): This the amount (in Euros) actually taken from the wallet. It takes on values of 0 to 5, by units of 0.5.

We also define deviation as the spread between DWA, EWA, OWA and AA. Namely, we have the Declared Deviation Amount (**DDA**=DWA-AA), the Expected Deviation Amount (**EDA**=EWA-AA) and the Observed Deviation Amount (**ODA**=OWA-AA).

Last, as participants do not receive the same AA, we need to control for the honesty behavior that the draw allows them to reveal. We cannot observe dishonesty for a 10 black-ball draw even though this participant may have been dishonest for a different AA. To control for the effect of a more- or less-fortunate draw, we define the mean of the Possible Deviation Amount ($\overline{\boldsymbol{PDA}}$) as the average of the positive spread that can be observed given the AA. For instance, a participant receiving the most-favorable draw cannot be identified as dishonest: $AA = 5$ corresponds to $\overline{PDA} = 0$. When

$AA = 2.5$, a participant who decides to be dishonest can take 0.5, 1, 1.5, 2 or 2.5€ more than the AA. If the AA does not affect the intensive margin of dishonesty, we should observe, on average, $ODA = \overline{PDA} = \frac{(0.5+1+1.5+2+2.5)}{5} = 1.5$ for dishonest participants. Our measure of the intensive margin of dishonesty will be $ODA - \overline{PDA}$ for those who are dishonest ($ODA > 0$).

Table 2 summarizes the mean values of AA, DDA, EDA and ODA across our different treatments. The detailed analyses appear in the next sections.

Table 2: The overall results across the two treatments

Treatment	N	AA		DDA		EDA		ODA	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Believe+Behave	88	2.47	1.61	0.16€	0.82	0.18€	0.99	0.31€	0.94
Behave+Believe	77	2.53	1.57	0.10€	0.56	0.20€	0.76	0.14€	1.42
Total	165	2.5	1.59	0.13€	0.71	0.19€	0.88	0.23€	1.19

Notes: AA = Allowed Amount, DDA = Declared Deviation Amount, EDA = Expected Deviation Amount, and ODA = Observed Deviation Amount.

4.2 Honesty behaviors

Result 1 *Participants are significantly dishonest, in particular when they receive an unfavorable draw.*

We find that participants are significantly dishonest: they took 0.23€ more than the allowed amount (AA): $ODA = 0.23 > 0$ (p-value=0.003), see Table 2 (Section 4.1). 21% of participants deviate ($ODA > 0$). This percentage falls with the AA (i.e. as the $\overline{PDA}$ increases): see Model (3) of Table 3. Among dishonest participants, the average $\overline{PDA}$ (1.82€) is not significantly different from their average ODA (1.74€), p-value=0.25. The ODA rises uniformly with the $\overline{PDA}$. In Model (2) of Table 3, we look at the ODA of participants who deviate. The estimated coefficient on $\overline{PDA}$ (0.95**) is not significantly different from 1. In other words, as the $\overline{PDA}$ rises by 1€ , the degree of dishonesty increases by 1€. An unfavorable draw increases dishonest behavior at the extensive margin (i.e. probability to be dishonest) but have no effect at the intensive margin (i.e. the degree of the dishonesty).

Result 2 *More a participant believes that others are dishonest, more she is dishonest, whatever the task order.*

On average, participants who first consider others' honesty (the *Believe+Behave* treatment) took 0.31€ more than the allowed amount (AA) while those who played the honesty task

first (*Behave+Believe*) took on average 0.14€ more: see Table 2 (Section 4.1). However, this difference is not significant (Models (1), (2) and (3) in Table 3). As the main consequence, forcing participants to consider others' honesty first does not then affect their own honesty.

In Models (1), (2) and (3) of Table 3, we add the individual beliefs in others' honesty. This latter is the average EDA for each participant ($\overline{EDA}_i$) who gave answers regarding the 11 possible draws. Participants who believe that others are dishonest are significantly more dishonest themselves (with a coefficient of 0.57**): those who believe that others will take one additional Euro took on average 0.47€ more than the AA ($sd = 0.16$). This is supported by greater dishonesty at the extensive margin (Model (3) in Table 3).

We add an interaction between $\overline{EDA}_i$ and the *Believe+Behave* dummy. Beliefs in others' honesty turn out not to have a differential impact on honesty when participants are primed to first think about others. However, it is possible that our treatments are not able to totally capture this kind of effect. If participants always adjust their behavior according to their beliefs about others, then there will be no additional effect from emphasizing other people's behavior.

Result 3 *Men are more dishonest than women. This is neither explained by an increase of the probability of being dishonest nor an inrease of the degree of dishonesty but by a combination of both.*

The difference between the ODA's in the *Believe+Behave* and *Behave+Believe* treatments mainly reflects the gender difference in the two samples, with respectively 46% and 34% men (see Table 1 in Section 3) and men being significantly more dishonest (ODA=0.46€) than women (ODA=0.08€) (p-value=0.003, Model(1) of Table 3). The gender effect is positive but insignificant at both the intensive and extensive margins (Models (2) and (3) of Table 3); this may reflect the relatively few individuals who cheat in Model (2). We thus conclude that the significant difference in Model (1) comes from the combination of both.

Table 3: Own honesty behavior

	(1)	(2)	(3)
	ODA	$ODA \mid ODA > 0$	$ODA > 0 \mid AA \neq 5$
$\overline{PDA}$	0.51***	0.95***	0.64**
	(0.10)	(0.34)	(0.30)
Treatment = *Believe+Behave*	0.22	-0.11	0.51

	(1)	(2)	(3)
	ODA	$ODA \mid ODA > 0$	$ODA > 0 \mid AA \neq 5$
	(0.19)	(0.67)	(0.53)
Male	0.36**	0.29	0.57
	(0.17)	(0.47)	(0.42)
Age	0.08	-0.06	0.28*
	(0.06)	(0.13)	(0.15)
$\overline{EDA_i}$	0.57**	0.44	1.25**
	(0.25)	(0.47)	(0.61)
$\overline{EDA_i}\times$ Treatment = *Believe+Behave*	-0.13	-0.23	0.61
	(0.32)	(0.79)	(0.82)
Constant	-2.44**	0.99	-8.48***
	(1.23)	(2.85)	(3.02)
N	165	35	150
No. of subjects	165	35	150
Method	OLS	OLS	Logit
Sample	All	$ODA > 0$	$AA \neq 5$

Notes: Significance : $*$ = 10% $**$ = 5% $***$ = 1%. Standard errors are in parentheses. The reference treatment is *Behave+Believe*. Models (1) and (2) refer to Ordinary Least Squares estimation, and Model (3) to Logistic estimation. AA= Allowed Amount, $\overline{EDA_i}$ = average Expected Deviation Amount for each participant, ODA = Observed Deviation Amount and $\overline{PDA}$= average Possible Deviation Amount (which falls in AA).

4.3 Beliefs in others' honesty

Result 4 *Participants significantly believe that others are dishonest. Beliefs in others' honesty are lower for unfavorable draws.*

Participants significantly believe that others are dishonest (i.e. $EDA = 0.19 > 0$, p-value$<$0.001), see Table 2 in Section 4.1. 21.9% of participants believe that others are dishonest ($EDA > 0$). As for participants' own honesty, this percentage falls with AA (and so rises with $\overline{PDA}$): see Model (6) of Table 4. Among these 21.9% of participants, the average EDA=1.25 is significanty lower than the $\overline{PDA}$=1.95 (p-value$<$0.001). Unlike for participants' own honesty, the EDA grows slower than the $\overline{PDA}$. In Model (5) of Table 4, we look at the EDA of participants who believe that others deviate: the coefficient on $\overline{PDA}$ (0.54***) is significantly below 1, so that as $\overline{PDA}$ rises by 1€, the EDA only increases

by one half. Beliefs in others' honesty therefore fall for unfavorable draws at the extensive margin, but rise at the intensive margin. Overall, an unfavorable draw reduces beliefs in others' honesty (Model (4) of Table 4).

Result 5 *Task ordering modifies beliefs in others' honesty. The percentage of participants expecting a deviation is greater in the Believe+Behave treatment. Overall, this effect is offset by an increase in beliefs regarding the degree of dishonesty for participants in the Behave+Believe treatment. Only participants who first face the honesty task take into account their own degree of dishonesty when assessing the degree of others' dishonesty.*

Participants in the *Believe+Behave* treatment expect an average deviation of 0.18€, which is not significantly different (p-value=0.124) from the figure in the *Behave+Believe* treatment: 0.20€, see Table 2 (Section 4.1). Facing own honesty first does not then seem to change beliefs about others' honesty. However, we find opposing effects when we separate out the effect of task ordering on the intensive and extensive margins. At the extensive margin, 25.7% of the participants in the *Believe+Behave* treatment believe in others' dishonesty (EDA>0) while this figure is 17.6% in the *Behave+Believe* treatment (the *Believe+Behave* dummy in Model (6) of Table 4 is significant at the 3% level). But at the intensive margin, participants' EDA rises in their own ODA only for those who face their own honesty first ($0.16 - 0.23$ is not significantly different from 0 in Model (5) of Table 4), which offsets the effect of the difference between the two treatments regarding the extensive margin of others' dishonesty.

Result 6 *Self-projection regarding beliefs in others' honesty (i.e. beliefs based on own behavior) is dominated by a mirroring effect (i.e. own behavior based on beliefs about others' behavior).*

Result 2 and *Result 5* reveal a strong relationship between honesty beliefs and honesty behaviors. However, with two-way causality and without a satisfactory instrumental variable we cannot determine in which way the causality runs. We address this issue by considering the gender difference in beliefs and own honesty. While men are significantly less honest than women, independently of their beliefs (*Result 3*), there is no difference between men and women in beliefs about others' honesty (Model (4), Table 4). In Model (4), we could argue that the effect is captured by the ODA, as men have higher ODA. We thus run an additional regression removing the ODA variable. We find no gender effect on beliefs in others' honesty (EDA), see Model (4 bis) of Table 4. This suggests that individuals adjust their honesty behavior according to the honesty they expect from others. Were the reverse causality to dominate (i.e. beliefs are based on own behavior), we should have found a

significantly higher EDA for men than for women.

Table 4: Beliefs in others' honesty

	(4)	(4 bis)	(5)	(6)
	EDA	EDA	$EDA \vert EDA > 0$	$EDA > 0 \vert AA \neq 5$
$\overline{PDA}$	0.34***	0.34***	0.54***	0.90***
	(0.02)	(0.02)	(0.07)	(0.09)
Treatment = *Believe+Behave*	-0.05	-0.02	-0.01	0.40***
	(0.04)	(0.04)	(0.11)	(0.13)
Male	0.00	0.04	-0.05	0.1
	(0.04)	(0.04)	(0.11)	(0.13)
Age	0.00	0.00	0.02	-0.06
	(0.01)	(0.01)	(0.05)	(0.05)
ODA	0.09***	-	0.16***	0.16**
	(0.02)	-	(0.04)	(0.07)
ODA × Treatment = *Believe+Behave*	0.03	-	-0.23***	0.07
	(0.04)	-	(0.07)	(0.1)
Constant	-0.22	-0.35	-0.1	-1.95**
	(0.29)	(0.29)	(0.92)	(0.98)
N	1815	1815	398	1650
No. of subjects	165	165	96	165
Method	OLS	OLS	OLS	Logit
Sample	All	All	$EDA > 0$	$AA \neq 5$

Notes: Significance : $*$ = 10% $**$ = 5% $***$ = 1%. Standard errors are in parentheses. The reference treatment is *Behave+Believe*. Models (4) and (5) refer to Ordinary Least Squares estimation, and Model (6) to Logistic estimation. AA= Allowed Amount, EDA = Expected Deviation Amount, ODA = Observed Deviation Amount and $\overline{PDA}$= average Possible Deviation Amount (which falls in AA).

4.4 Declared and observed behaviors

Result 7 *There is hypothetical bias for honest and dishonest participants. Honest participants overestimate their dishonesty for unfavorable draws, while dishonest participants underestimate their dishonesty for favorable draws. In addition, participants who said they would be dishonest took more from the wallet than the amount they had declared.*

From the *Believe+Behave* treatment, we examine the spread between the participants' declared amount taken (DDA) and the amount actually taken (ODA), given the draw (AA). As a reminder, between the two tasks we ask participants the following hypothetical questions: *"How many Euros would you have taken if you had received this draw?"*. Participants in the *Believe&Behave* treatment have not yet carried out the honesty task, and do not know what the next step is in the experiment.

Table 5 summarizes the average AA, DDA and ODA figures given participants' observed deviations ($ODA > 0$ or $ODA \leq 0$) and their declared deviation ($DDA > 0$ or $DDA \leq 0$). We here only use information on the participant's DDA figure that matches the actual draw they received during the honesty task.

In the *Believe+Behave* treatment, participants declared a deviation with respect to their draw received (DDA) of 0.18€ (see the bottom right of Table 5), while the actual observed deviation (ODA) was 0.31€. This difference is not statistically significant (p-value=0.458). The percentage of participants who actually deviated (22.7%) is also not significantly different from the percentage who said that they would deviate (23.9%). However, only half of participants who do deviate (ODA>0) said that they would do so (DDA>0) and the other half did not (DDA $\leq$ 0). Honest participants overestimate their dishonesty for unfavorable draws: the average AA of those who correctly declare their honesty (2.72€) is higher than that of those who do not (1.91€; p-value=0.066). On the contrary, dishonest participants underestimate their dishonesty for favorable draws: the average AA of participants who correctly declare their dishonesty (1.91€) is lower than that of those who do not (2.50€; p-value=0.085). In addition, participants who declare that they would deviate underestimate their degree of dishonesty. On average, their DDA=1€ is lower than their ODA=2.2€ (p-value=0.02).

Table 5: Hypothetical biases analysis (*Believe+Behave* treatment)

	DDA$\leq$0	DDA>0	Total
$ODA \leq 0$	64.8% (N=57)	12.5% (N=11)	77.3% (N=68)
	ODA=-0.06 (0.34)	ODA=-0.09 (0.20)	ODA=-0.07 (0.32)
	DDA=-0.07 (0.35)	DDA=1.00 (1.10)	DDA= 0.10 (0.66)
	AA=2.72 (1.56)	AA=1.91 (1.92)	AA=2.59 (1.63)
$ODA > 0$	11.4% (N=10)	11.4% (N=10)	22.7% (N=20)
	ODA=1.00 (0.71)	ODA=2.20 (1.27)	ODA=1.60 (1.18)

	DDA≤0	DDA>0	Total
	DDA=-0.10 (0.32)	DDA=1.00 (0.71)	DDA=0.45 (0.78)
	AA=2.50 (1.20)	AA=1.65 (1.67)	AA=2.08 (1.48)
Total	76.1% (N=67)	23.9% (N=21)	100%(N=88)
	ODA=0.10 (0.56)	ODA=1.00 (1.46)	ODA= 0.31 (0.94)
	DDA=-0.07 (0.34)	DDA=1.00 (0.91)	DDA=0.18 (0.70)
	AA=2.69 (1.50)	AA=1.79 (1.76)	AA=2.47 (1.61)

Notes: The ODA, DDA and AA figures are means with standard errors in parentheses. The analysis is based on the *Believe+Behave* sample. AA= Allowed Amount, DDA = Declared Deviation Amount for the draw received in the honesty task, and ODA = Observed Deviation Amount.

Result 8 *During the ex-post survey, participants hid being dishonest (at the extensive margin) and understated how dishonest they are (at the intensive margin).*

Table 6 summarizes the average AA, DDA and ODA figures given participants' observed deviations ($ODA > 0$ or $ODA \leq 0$) and their declared deviation ($DDA > 0$ or $DDA \leq 0$). We here only use information on the participant's DDA figure that matches the actual draw they received during the honesty task.

In the *Behave+Believe* treatment, the declared and observed deviations are consistent. The average ODA is equal to the average DDA (0.14€, at the bottom right of Table 6). Participants who do not deviate (80.5%) have no interest in misreporting their behavior: only 3 of the 68 honest participants (ODA≤0) have a DDA>0. Only half of the dishonest (ODA>0) declared being so (DDA>0). Among the participants who declared their dishonesty (DDA>0 and ODA>0), the degree of dishonesty was understated: their DDA=1.21€ is lower than their ODA=2.64€ (p-value=0.10).

Table 6: Self-reporting (*Behave+Believe* treatment)

	DDA≤0	DDA>0	Total
	76.6% (N=59)	3.9% (N=3)	80.5% (N=62)
	ODA=-0.30 (0.95)	ODA=-0.33 (0.58)	ODA=-0.30 (0.93)
ODA≤0	DDA=-0.04 (0.23)	DDA=1.67 (2.02)	DDA=0.04 (0.57)
	AA=2.83 (0.81)	AA=1.67 (0.76)	AA=2.77 (1.53)
	10.4% (N=8)	9.1% (N=7)	19.5% (N=15)
	ODA=1.31 (1.28)	ODA=2.64 (1.95)	ODA=1.93 (0.93)
ODA>0			

	DDA≤0	DDA>0	Total
	DDA=0 (0)	DDA=1.21 (1.47)	DDA=0.57 (1.15)
	AA=1.94 (1.64)	AA=1.07 (0.89)	AA=1.53 (1.37)
Total	87.0% (N=67)	13.0% (N=10)	100% (N=77)
	ODA=-0.10 (1.11)	ODA=1.75 (2.16)	ODA=0.14 (1.42)
	DDA=-0.04 (0.22)	DDA=1.35 (1.55)	DDA=0.14 (0.74)
	AA=2.72 (1.57)	AA=1.25 (0.86)	AA=2.53 (1.57)

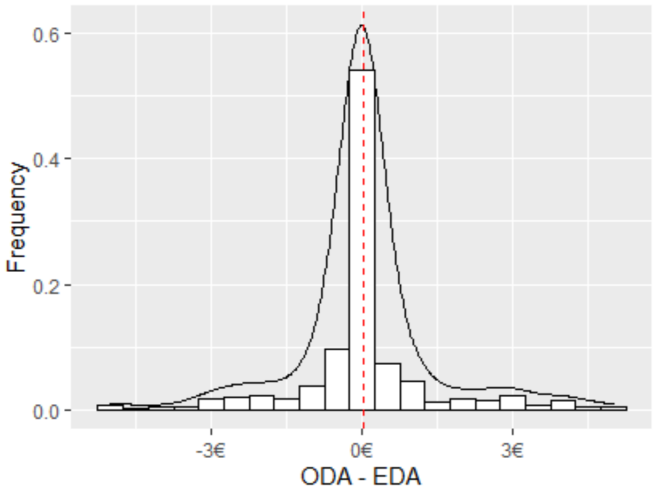
Notes: The ODA, DDA and AA figures are means with standard errors in parentheses. The analysis is based on the *Behave+Believe* sample. AA= Allowed Amount, DDA = Declared Deviation Amount for the draw received in the honesty task, and ODA = Observed Deviation Amount.

4.5 Matching beliefs with behaviors about honesty

Result 9 *Participants are less honest than others think they are. There is no significant difference at the extensive margin, but honesty is overestimated (underestimated) at the intensive margin for favorable (unfavorable) draws.*

We consider the spread between the ODA and the EDA for each pair of participants to address our main research question: Are we more honest than others think we are? The honesty behavior of an individual i, with a certain AA draw, will be matched to the beliefs of all of the other j, $j \neq i$, participants of how an individual would behave when faced with that AA draw. This produces 27 060 matches of what an individual actually does to what everyone else expected them to do. Figure 1 shows the distribution of the $ODA - EDA$ spread; the mean figure here is 0.04€ (sd=1.39), which is significant (p-value<0.01). In other words, participants significantly underestimate others' dishonesty.

Figure 1: The empirical distribution of the ODA-EDA pairs



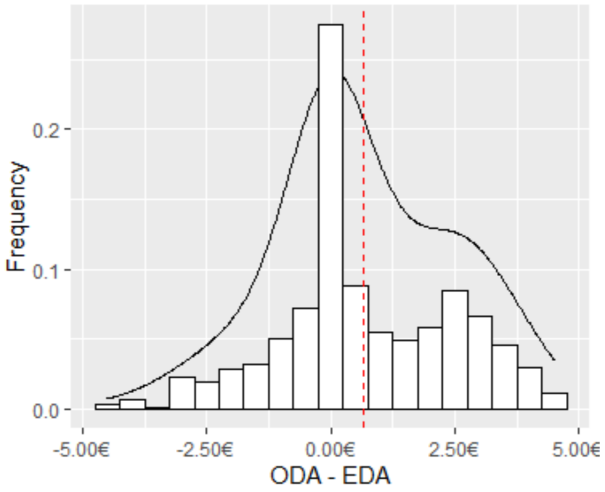
Notes: The dashed line is the sample mean. AA= Allowed Amount, EDA = Expected Deviation Amount and ODA = Observed Deviation Amount.

54% of the pairs match (i.e. ODA-EDA=0). Of these matching pairs, 97% correspond to ODA=EDA=0 (participants behave according to the rule and others believe that they will do so). This 97% figure is unsurprising, as the number of combinations of the answers for participants who do not believe in others' honesty ($EDA \neq 0$, 10 possible answers) and the actual behavior of participants who do not follow the rule ($ODA \neq 0$, 10 possible behaviors) is exponentially larger ($= 10^2$) than the number of combinations when ODA=0 or EDA=0 ($= 10$). The probability that $ODA = EDA$ in this case does not provide useful information.

We thus consider the intensive margin of honesty by modeling the $ODA - EDA$ spread when participants correctly anticipated dishonesty ($EDA >$ and $ODA > 0$). We also look at the extensive margin of honesty: the difference between the frequencies of $ODA > 0$ and $EDA > 0$.

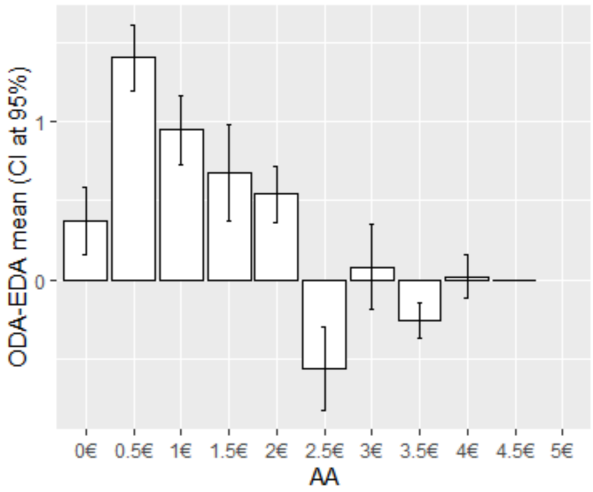
Figure 2 shows the distribution of the $ODA - EDA$ spread when participants correctly anticipated dishonesty ($EDA > 0$ and $ODA > 0$); the mean figure here is 0.67€ (sd=1.77), which is significant (p-value<0.01). Participants then underestimate the degree of others' dishonesty. In particular, as shown in Figure 3, participants' degree of dishonesty is overestimated (underestimated) for favorable (unfavorable) draws. Participants exhibit the same degree of dishonesty with respect to the allowed amount (*Result 1*) while others believe that they will become more dishonest at the intensive margin as the AA increases (*Result 4*).

Figure 2: The empirical distribution of the ODA-EDA pairs | ODA>0 and EDA >0



Notes: The dashed line is the sample mean. AA = Allowed Amount, EDA = Expected Deviation Amount and ODA = Observed Deviation Amount.

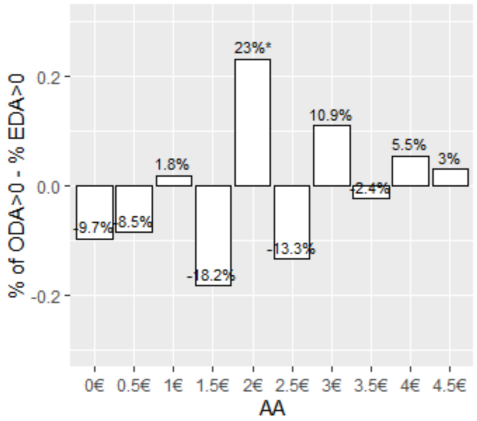
Figure 3: The means of the ODA-EDA pairs by AA | ODA>0 and EDA >0



Notes: AA = Allowed Amount, EDA = Expected Deviation Amount and ODA = Observed Deviation Amount. The vertical lines are the 95% confidence intervals.

At the extensive margin, we find no difference between the percentage of participants who expect deviation ($EDA > 0$ for 21.9% of participants) and the percentage of participants who do actually deviate ($ODA >$ for 21.1% of participants), regardless of the AA (Figure 4).

Figure 4: The difference (in pts) between % ODA>0 and % EDA>0



Notes: The significance refers to a χ^2 test between % ODA>0 and % EDA>0 (27.2%): $*$ = 10% $**$ = 5% $***$ = 1%. AA = Allowed Amount, EDA = Expected Deviation Amount and ODA = Observed Deviation Amount.

Model (7) in Table 7 shows the OLS estimation results for the $ODA - EDA$ spread including individual random effects. We include the characteristics of the participants on both sides of the pair. We drop participant age from the analysis since, as in the other regressions, it was never significant. This is probably due to the only limited age variation in our sample (see Table 1).

Gender is significantly correlated with the spread between the individual's beliefs and others' actual honesty. In particular, the overestimation of honesty ($ODA - EDA > 0$) is greater when we only consider the honesty behavior of men (the $Gender = Male|ODA$ coefficient is significantly positive in Table 7). Men are significantly more dishonest than women (*Result 3*). There is no difference in the fit between honesty and beliefs in others' honesty between men and women (the $Gender = Male|EDA$ coefficient is insignificant. This result is in line with *Result 6*. Participants adjust their behavior according to their beliefs in others' honesty and not their beliefs regarding their own honesty.

Table 7: Honesty beliefs and behaviors

	(7)
	$ODA - EDA$
AA	0.08***
	(0.01)
AA × ODA>0	-0.38***
	(0.01)
ODA>0	2.50***
	(0.03)

	(7)
	$ODA - EDA$
Male | ODA	0.16***
	(0.02)
Male | EDA	-0.03
	(0.08)
Constant	-0.58***
	(0.06)
N	27060
No. of subjects	165
Method	OLS incl. Ind. Random effects
Sample	All

Notes: Significance : $* = 10\%$ $** = 5\%$ $*** = 1\%$. These are Ordinary Least Squares estimates. The regression includes individual random effects. The honesty behavior of an individual i, with a certain AA draw, is matched to the beliefs of all of the other j individuals. We have multiple beliefs in others' honesty for each participant j, corresponding to the different draws received by the i participants in the honesty task. AA= Allowed Amount, EDA = Estimated Deviation Amount and ODA= Observed Deviation Amount. ODA>0 is a dummy for participant i being dishonest. Male|ODA, Male|EDA and Male|EDA,ODA are dummy variables for i, j, and both i and j being men.

5 DISCUSSION AND CONCLUSION

This original experiment examines the relationship between individuals' honesty and their beliefs in others' honesty under more-or less- favorable conditions. Honesty is viewed as compliance with a given rule where it is also possible to cheat only partially. Honesty is not monitored, there are no social interactions, and all decisions are anonymous. Participants are randomly assigned a more- or less-fortunate initial state that determines their earnings if they follow the rule. The experiment aims to analyze three key characteristics: honesty, beliefs in others' honesty, and the impact of more- or less-fortunate initial situations.

Similarly to Yaniv & Siniver (2016), we find that participants significantly cheat according to an objective rule in an unmonitored environment. In addition, participants significantly believe that others cheat. However, participants are less honest than others think they are.

Unfavorable conditions increase both own dishonesty and beliefs in others' dishonesty (in line with Houser et al. (2012) and Galeotti et al. (2017)), and the spread between both is explained by the difference in what participants think about the effects of more- or less-favorable draws and how participants actually act with these draws. As in Ermisch

et al. (2009), we find that participants are more likely to believe that others will not cheat in "comfortable" situations, where "comfortable" situations refers here to more fortunate draws.

Our results demonstrate again the importance of taking into account the difference between honest and dishonest earnings (as in Gibson et al. (2013), Gneezy et al. (2013) and Gneezy et al. (2018)). A greater difference between honest earnings and possible dishonest earnings increases dishonesty and beliefs in others' dishonesty at the extensive margin (i.e. the probability of being dishonest). However, the decision to be dishonest is conditioned by a more- or less-favorable environment only at the extensive margin. Once the decision to be dishonest has been taken, the degree of dishonesty is unchanged, independently of the environment. Therefore, while the frequency of dishonest behaviors is correctly anticipated for the different AA, the degree of dishonesty is underestimated (overestimated) in unfavorable (favorable) conditions.

Last, honesty and beliefs in others' honesty are strongly correlated. In contrast to Hugh-Jones (2016), this result suggests that participants believe that others are dishonest and then become dishonest in turn. As in Robert & Arnab (2012), it seems that dishonesty is contagious. An increase of beliefs in others' honesty reduces dishonest behavior. While there is probably two-way causality between both, we control for the difference in gender behavior. Men are significantly more dishonest but have the same level of beliefs in others' dishonesty as women. Thus, any self-projection on beliefs in others' honesty (i.e. beliefs based on own behavior) is dominated by a mirroring effect (i.e. own behavior based on beliefs about others' behavior).

While these results are consistent with a number of findings in the existing literature, they also raise questions about the interest of providing information about others' honesty. Rauhut (2013) shows that providing information about others' dishonesty leads participants who underestimated the frequency of dishonest behaviors to become less honest themselves. In addition, Maux et al. (2021) find that the beliefs of participants who received information converge to the observed level of dishonesty, thus reducing the importance of beliefs on decision-making. In a context in which people are less honest than others think, the adequacy between honesty and beliefs comes at the risk of increasing overall dishonesty.

Beyond the frequency of dishonest behavior, what about the extent of the dishonesty ? Our experimental results show that the frequency of dishonest behaviors is correctly predicted by the participants, without any information. It is the degree of dishonesty which is

overall underestimated (with opposite effects depending on the more- or less-favorable environment).

Last, and at the difference of the experiment carried out by Robert & Arnab (2012) (a deception game), the experimental design does not include here information and monetary interaction between participants. Without information, beliefs in others' honesty make "supposed" others' dishonesty contagious too, even when the dishonesty of others does not have any consequences on participants' own earnings. What might happen if honesty and beliefs in others' honesty includes a group dimension, where the dishonest behaviors of some reduce the honest earnings of others? Berg et al. (1995) and Glaeser et al. (1996, 2000) find that participants are more likely to trust and to be trusted in turn when they interact. Including a group dimension may thus increase individual honesty. However, without concrete information on others' actual behavior, beliefs become even more important. Those who believe that others are dishonest become less honest themselves. In the group context, the supposed dishonesty of others also reduces the individual's own honest earnings, which will increase the incentives to be dishonest. The beliefs about the honesty of others can allow to justify individuals own turpitude. The level of the overall (dis)honesty depends on the strength of this specular reasoning.

References

Abeler, J., Nosenzo, D. & Raymond, C. (2019), 'Preferences for truth-telling', *Econometrica* **87**(4), 1115–1153.

Andersen, S., Gneezy, U., Kajackaite, A. & Marx, J. (2018), 'Allowing for reflection time does not change behavior in dictator and cheating games', *Journal of Economic Behavior and Organization* **145**, 24–33.

Balasubramanian, P., Bennett, V. M. & Pierce, L. (2017), 'The wages of dishonesty: The supply of cheating under high-powered incentives', *Journal of Economic Behavior and Organization* **137**(C), 428–444.

Becker, G. S. (1968), 'Crime and punishment: An economic approach', *Journal of Political Economy* **76**(2), 169–217.

Berg, J., Dickhaut, J. & McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122 – 142.

Cohn, A., Marechal, M. A. & Noll, T. (2015), 'Bad boys: How criminal identity salience affects rule violation', *Review of Economic Studies* **82**(4), 1289–1308.

Cohn, A., Maréchal, M. A., Tannenbaum, D. & Zünd, C. L. (2019), 'Civic honesty around the globe', *Science* **365**(6448), 70–73.

Ermisch, J., Gambetta, D., Laurie, H., Siedler, T. & Noah, U. S. C. (2009), 'Measuring people's trust', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(4), 749–769.

Fischbacher, U. & Follmi-Heusi, F. (2013), 'Lies in disguise - an experimental study on cheating', *Journal of the European Economic Association* **11**(3), 525–547.

Galeotti, F., Kline, R. & Orsini, R. (2017), 'When foul play seems fair: Exploring the link between just deserts and honesty', *Journal of Economic Behavior and Organization* **142**, 451 – 467.

Gibson, R., Tanner, C. & Wagner, A. F. (2013), 'Preferences for truthfulness: Heterogeneity among and within individuals', *American Economic Review* **103**(1), 532–48.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. & Soutter, C. L. (2000), 'Measuring trust', *Quarterly Journal of Economics* **115**(3), 811–846.

Glaeser, E. L., Sacerdote, B. & Scheinkman, J. (1996), 'Crime and social interactions', *Quarterly Journal of Economics* **111**(2), 507–548.

Gneezy, U., Kajackaite, A. & Sobel, J. (2018), 'Lying aversion and the size of the lie', *American Economic Review* **108**(2), 419–53.

Gneezy, U., Rockenbach, B. & Serra-Garcia, M. (2013), 'Measuring lying aversion', *Journal of Economic Behavior and Organization* **93**(C), 293–300.

Grolleau, G., G. Kocher, M. & Sutan, A. (2016), 'Cheating and loss aversion: Do people cheat more to avoid a loss?', *Management Science* **62**(12), 3428–3438.

Houser, D., Vetter, S. & Winter, J. (2012), 'Fairness and cheating', *European Economic Review* **56**(8), 1645 – 1655.

Hugh-Jones, D. (2016), 'Honesty, beliefs about honesty, and economic growth in 15 countries', *Journal of Economic Behavior and Organization* **127**(C), 99–114.

Kajackaite, A. & Gneezy, U. (2017), 'Incentives and cheating', *Games and Economic Behavior* **102**(C), 433–444.

Maux, B. L., Masclet, D. & Necker, S. (2021), 'Monetary incentives and the contagion of unethical behavior', *SSRN Electronic Journal* .

Mazar, N., Amir, O. & Ariely, D. (2008), 'The dishonesty of honest people: A theory of self-concept maintenance', *Journal of Marketing Research* **45**(6), 633–644.

Nagin, D. & Pogarsky, G. (2003), 'An experimental investigation of deterrence: cheating, self-serving bias, and impulsivity', *Criminology* **41**, 167–194.

Pruckner, G. & Sausgruber, R. (2013), 'Honesty on the streets: a field study on newspaper purchasing', *Journal of the European Economic Association* **11**(3), 661–679.

Rauhut, H. (2013), 'Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments', *PLoS ONE* **8**(11), e77878.

Robert, I. & Arnab, M. (2012), 'Is dishonesty contagious?', *Economic Inquiry* **51**(1), 722–734.

Yaniv, G. & Siniver, E. (2016), 'The (honest) truth about rational dishonesty', *Journal of Economic Psychology* **53**, 131 – 140.