Appendix

Appendix A1: Additional Results of the Lab Experiment

Figure A1.1 Distribution of reported sums of rolling two fair dice in Round 1 (truthful distribution depicted as a reference)

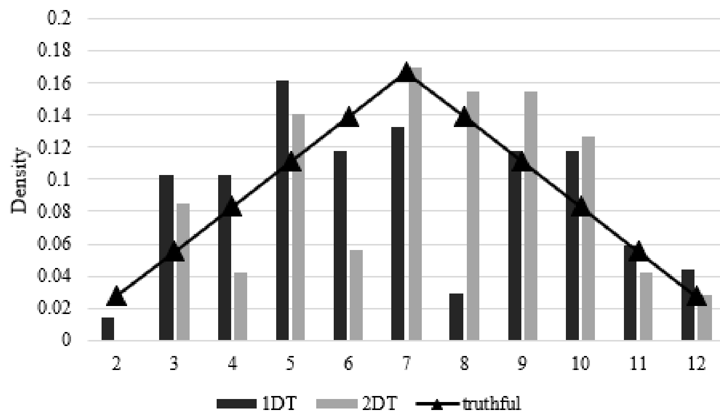
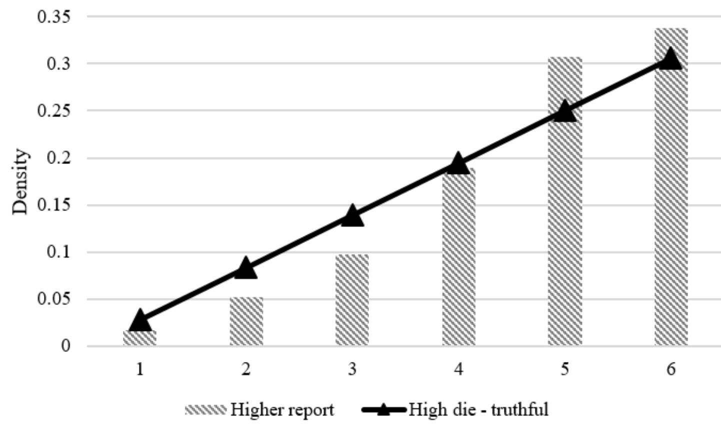
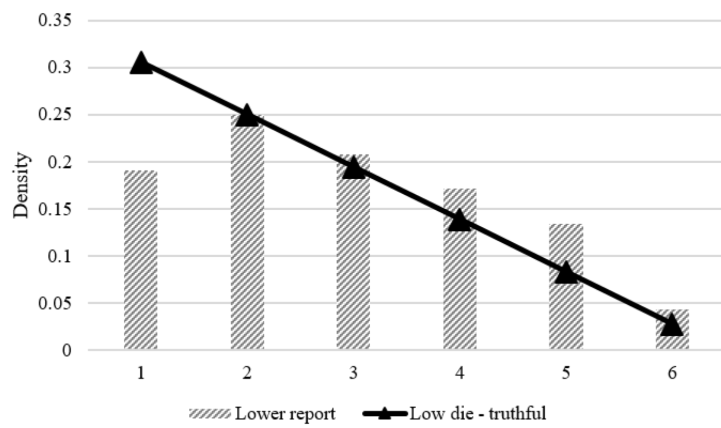


Figure A1.2 Distribution of lower and higher reports in rounds 2–10 of 2DT (truthful distribution depicted as a reference)

(a) Distribution of higher reports in rounds 2–10



(b) Distribution of lower reports in rounds 2–10



Model specifications for Table 2

In Table 2, we analyze the effect of earlier rounds on reports. For this purpose, we estimate the following equation:

$$y_{i,r} = \alpha_i + \gamma y_{i,r-1} + x'_{i,r} + v_{it}$$

where $i = 1,...,N$ individuals in $r = 1,...,10$ rounds. The variable $y_{i,r}$ is the outcome in the current round and $y_{i,r-1}$ is its lag, which is the sum of both dice in models (1) and (2), the lower die report in models (3) and the higher die report in models (4). The remaining variables are an unobserved individual-specific time-invariant fixed effect α_i and the ir-th observation of the explanatory variables $x_{i,r}$, here a temporal trend in models (b). Since the error term is correlated with the lagged dependent variable by construction, we estimate this equation using system GMM (Arellano and Bond, 1991). Deeper lags of the dependent variable are used as lags. In the presented specification, we use three lagged levels as instruments ($r-2$, $r-3$, $r-4$). To correct for potential bias in small samples, we further employ Windmeijer correction (Windmeijer, 2005). As indicated by the reported Arellano-Bond second-order serial correlation tests, we do not reject the null hypothesis of no higher-order serial correlation in first differences. We use the Stata command *xtabond2* for implementation (Roodman, 2009).

Appendix A2: Detailed Procedures and Results of the Lab-In-The-Field Experiment

Experimental Procedure

The experiment was conducted on two consecutive days. Students were brought from their classroom to another room where computers were set up for the experiment. The workspaces were divided by screens such that students could not observe each other during the experiment. We minimized information transmission between students by conducting the experiment during lessons, with a maximum of two groups per class to prevent the students from talking to each other after completing the session. Upon arrival in the room, participants were randomly assigned to 1DT or 2DT, resulting in group sizes of 56 and 54 participants, respectively. Students completed the same die-roll tasks as in the lab experiment over ten rounds. The only distinctive element was related to payoffs, which were determined as follows: (report-2)*€0.50. We designed lower expected payoffs in the lab-in-the-field experiment to account for the lower opportunity cost of adolescents. The 110 participants (45% female, and average age 15.95, balanced over treatments) received all instructions on-screen using oTree software (Chen et al., 2016). The average payment per student was €2.8 and was paid out at the end of the school day. Each session lasted 10 minutes.

Results regarding Hypothesis 2

- Average outcome in 1DT (7.7875) are shifted significantly from the distribution expected under truth-telling ($p<0.001$, KS; $p<0.001$, WSR).
- Average outcome in 2DT (7.6852) are shifted significantly from the distribution expected under truth-telling ($p<0.001$, KS; $p<0.001$, WSR).
- No difference in the level of lying between 1DT and 2DT ($p=0.823$, MW test).
- No difference in the fraction of *likely liars* between 1DT and 2DT ($p=0.648$, CT).

Results regarding Hypothesis 1

- Lying about the right and left dice is not significantly different in the first round (3.963 vs. 3.7963, $p=0.477$, WSR) nor in the nine later rounds (3.784 vs. 3.893, $p=0.368$, WSR).
- No difference in the fraction of *likely liars* between right die and left die ($p=0.108$, MNT).
- Deviations from the expected value in Round 1 are greater for the low reports than for the high reports, but this difference is not significant (0.2315 vs. 0.5278, $p=0.079$, TT). The same holds for the subsequent nine rounds (0.2706 vs. 0.4064, $p=0.068$, TT).

Figure A2.1 Distribution of reported sums of rolling two fair dice in the school sample (truthful distribution depicted as a reference)

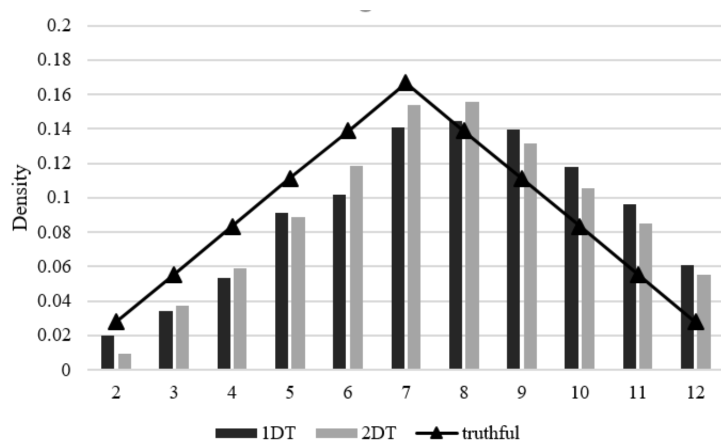


Table A2.1 Linear regression models of the level of lying in the school sample

Outcome variable: (Average report - 7)			
	(1)	(2)	(3)
2DT		-0.1023	-0.1108
		(0.2105)	(0.2145)
constant	0.7373***	0.7875***	0.7750***
	(0.1049)	(0.1475)	(0.1627)
Controls	No	No	Yes
N	110	110	110
R2	0	0.0022	0.0049

Standard errors in parentheses
Controls include gender and age
* p < 0.05, ** p < 0.01, *** p < 0.001

- Euclidean distance low die in Round 1: $d(r,t)_{low,1}$ = 0.2629 (confidence interval 0.1802–0.3456). Euclidean distance high die in Round 1: $d(r,t)_{high,1}$ = 0.1006 (confidence interval 0.0042–0.1970).

- In Rounds 2-10, Euclidean distance is greater for lower reports ($d(r,t)_{low,2-10}$ = 0.1176) than for higher reports ($d(r,t)_{high,2-10}$ = 0.0884), but estimates are included in each other's confidence intervals.

- No difference in the fraction of *likely liars* between lower and higher reports (p=0.285, MNT).

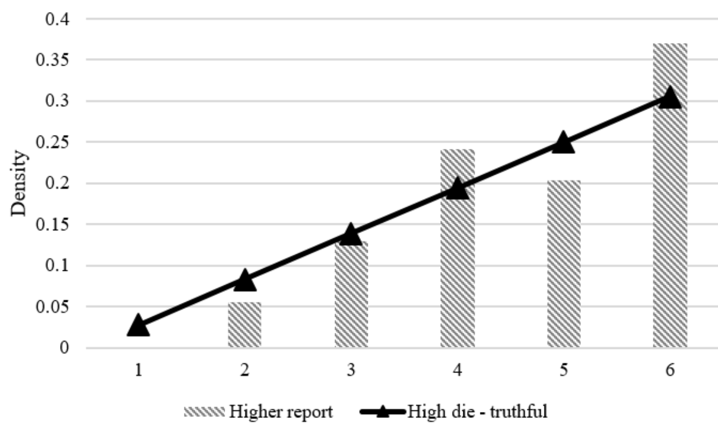
Table A2.2 Two-step system GMM with lagged levels (t-2, t-3 and t-4) of the dependent variable as instruments in the school sample

Outcome variable:	1DT diesum		2DT diesum		2DT lowdie		2DT highdie	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
l.outcome	0.0884	0.1164	-0.0312	-0.0039	-0.0265	-0.0053	-0.0247	-0.0278
	(0.0683)	(0.0681)	(0.0685)	(0.0642)	(0.0654)	(0.0625)	(0.0676)	(0.0665)
round		0.0239		0.0282		0.0196		0.0068
		(0.0420)		(0.0422)		(0.0298)		(0.0243)
constant	7.0358***	6.6754***	7.8512***	7.4618***	2.946***	2.756***	4.848***	4.8202***
	(0.5884)	(0.5654)	(0.5738)	(0.4826)	(0.2215)	(0.2173)	(0.3532)	(0.3413)
Arellano-Bond serial	-0.02	0.14	0.16	0.33	-0.74	-0.57	0.26	0.24
correlation test [p-value]	[0.987]	[0.396]	[0.877]	[0.739]	[0.461]	[0.570]	[0.794]	[0.811]
Instruments	30	31	30	31	30	31	30	31
Sample size	504	504	486	486	486	486	486	486

Robust standard errors using Windmeijer correction in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A2.2 Distribution of lower and higher reports in Round 1 of 2DT in the school sample (truthful distribution depicted as a reference)

(a) Distribution of higher reports in Round 1



(b) Distribution of lower reports in Round 1

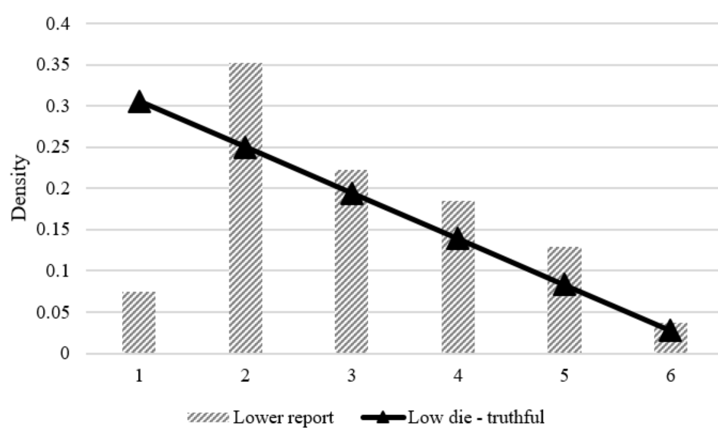
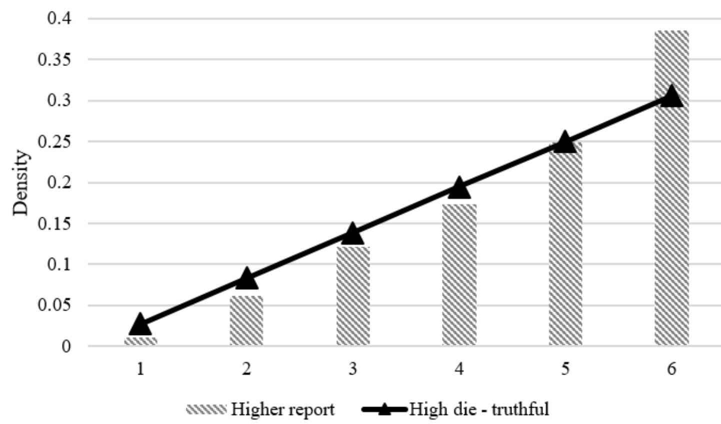
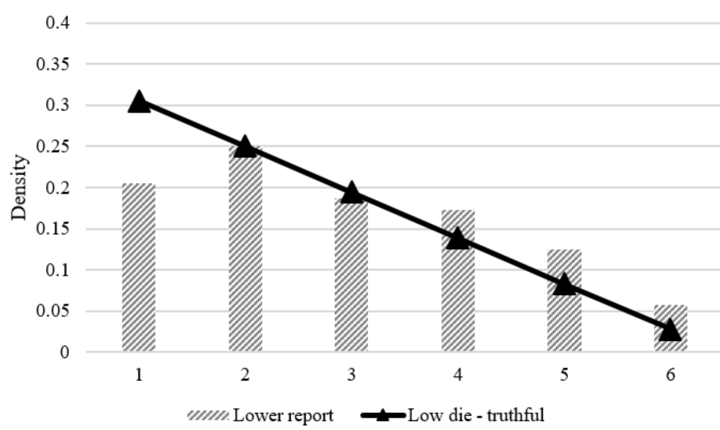


Figure A2.3 Distribution of lower and higher reports in rounds 2–10 of 2DT in the school sample (truthful distribution depicted as a reference)

(a) Distribution of higher report in rounds 2–10



(b) Distribution of lower report in rounds 2–10



Appendix B1: Experimental Instructions

The following shows the instructions for 1DT (2DT). Instructions were displayed on-screen, in English in the lab experiment and in German in the lab-in-the-field experiment.

Welcome
Thank you for participating in this experiment. All instructions will be given on-screen.
You are now taking part in an experiment in decision making. For showing up on time today, you will be paid €2. In addition, you can earn money with the decisions you make. Hence, it is important that you fully understand the instructions that follow. Please read them carefully.
Please enter the number of your computer so that your earnings at the end of the experiment can be matched to your computer cubicle. We use this number only to determine your payment. This means that the number will not be linked to your name for the data analysis. Your anonymity is therefore secured.
Please raise your hand if you have a question at any point of the experiment.

Instructions
In this experiment you will receive money based on the outcome of rolling two dice. You can find two dice in front of you on the table. You can inspect them now to see that they are regular 6-sided dice.
On the following screen you will be asked to roll both dice at the same time and to report the sum of eyes that you see. (On the following screen you will be asked to roll two dice, which we label "right die" and "left die", respectively, at the same time and to report the outcome of each die roll separately.) Your payment will be:
Sum of both dice*€0.50
(Outcome "right die"*€0.50 + Outcome "left die"*€0.50)
Example 1: If the sum of the two dice is 12 (If the outcome of both dice is 6), you will receive €6.
Example 2: If the sum of the two dice is 2 (If both dice show a 1), you will receive €1.

Repetition of task
Thank you for reporting the number of eyes of your dice. Let's do this again. The die roll task that you just completed will now be repeated for 9 more rounds. Your payoff for each round will be calculated as before:
Sum of both dice*€0.50
(Outcome "right die"*€0.50 + Outcome "left die"*€0.50)
At the end of the experiment, one of the 10 rounds will be randomly selected with equal probability for payment. Therefore, your final payment will be the payoff corresponding to the selected round.

Figure B.1 Reporting screen 1DT

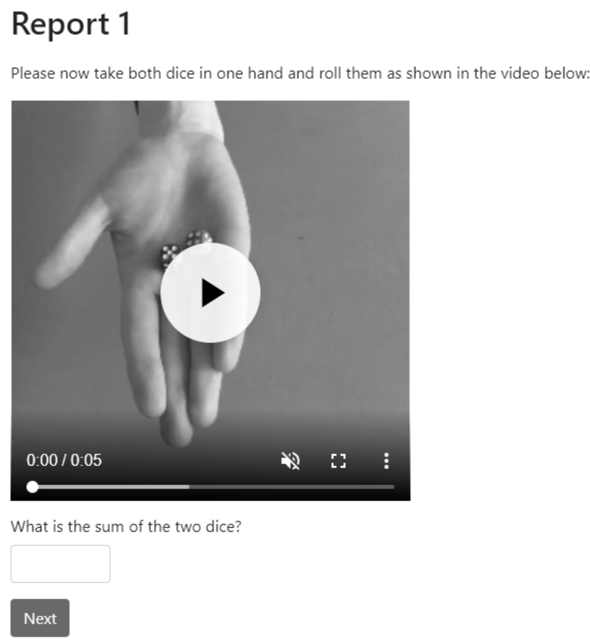
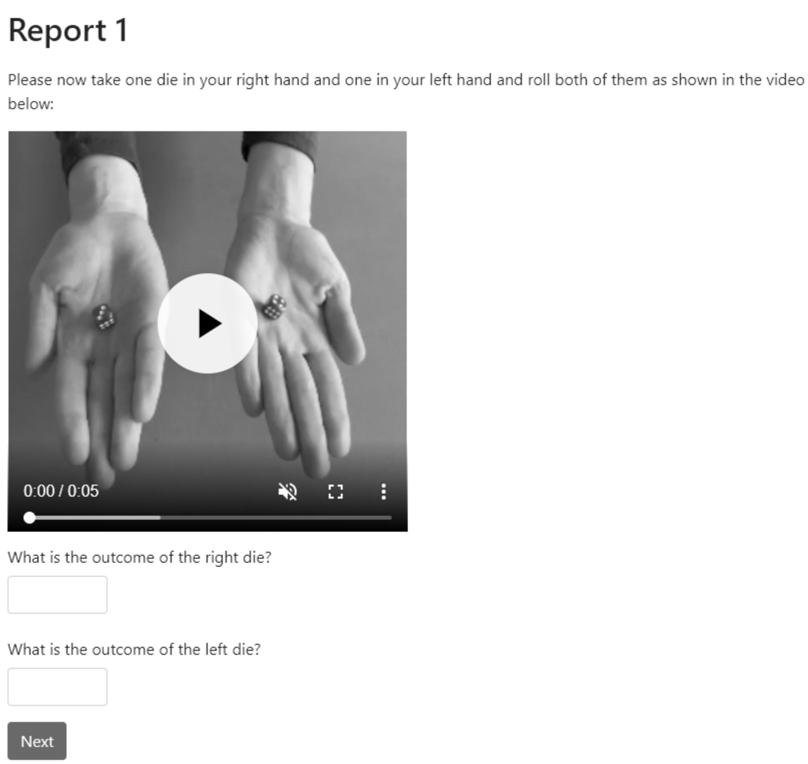


Figure B.2 Reporting screen 2DT



Appendix B2: Pros and Cons regarding the Incentive Scheme underlying the Task Repetition

As indicated in the paper (and reported in full detail in Appendix A1), the nine repetitions of the task were announced only after the first round, which makes this design element (slightly) deceptive.

In light of the fact that a non-deceptive solution does not exist that would allow us to compare moral balancing in lying in a one-shot situation to a repeated context, we considered that this design choice could be accepted as the scientific benefits outweigh possible drawbacks. For the sake of transparency regarding methodological standards of experimental economics, we discuss here the rationale for our decision:

One advantage of our design choice is that it allows us to collect more data from each participating subject. Most importantly, it also allows us to assess the first round as if it were a one-shot experiment. Thus, it allows us to assess a context in which only the presence of multiple simultaneous decisions can play a role in the decisions made. The obvious disadvantage is that participants may have become skeptical regarding the incentive scheme that applied to rounds 2-10, as they learned at the start of round 2 that round 1 is actually not paid for sure.