Journal of the Economic Science Association (2023) 9:34–50
https://doi.org/10.1007/s40881-022-00121-3

ORIGINAL PAPER

Lying in two dimensions

Diogo Geraldes[1] · **Franziska Heinicke**[2] · **Stephanie Rosenkranz**[3]

Received: 1 February 2021 / Revised: 6 November 2022 / Accepted: 7 November 2022 /
Published online: 29 November 2022

Abstract
The expanding literature on lying has exclusively considered lying behavior within a one-dimensional context. While this has been an important first step, many real-world contexts involve the possibility of simultaneously lying in more than one dimension (e.g., reporting one's income and expenses in a tax declaration). We experimentally investigate individual lying behavior in one- and two-dimensional contexts to understand how the multi-dimensionality of a decision affects lying behavior. Our paper provides the first evidence regarding the pure effect of dimensionality on lying behavior. Using a two-dimensional die-roll task, we show that participants distribute lies unevenly across dimensions, which results in greater over-reporting of the lower-outcome die.

Keywords Lying · Honesty · Multi-dimensional · Lab experiment · Lab-in-the-field experiment

JEL Classification C91 · C93 · D82 · H26

1 Introduction

In many real-world situations, people have the opportunity to lie for their own benefit. More often than not, these situations involve multi-dimensional decisions—that is, situations that involve the report of several outcomes in which the misreporting

✉ Diogo Geraldes
diogogeraldes@gmail.com

Franziska Heinicke
f.heinicke@uni-mannheim.de

Stephanie Rosenkranz
s.rosenkranz@uu.nl

1 School of Economics and Geary Institute, University College Dublin, Dublin, Ireland

2 Department of Economics, University of Mannheim, Mannheim, Germany

3 Utrecht School of Economics, Utrecht University, Utrecht, Netherlands

of each outcome yields self-benefits. In tax declarations, for example, people can under-report their income and over-report their expenses to pay less in taxes. Likewise, in the field of science, researchers have the option to manipulate the data collection and the statistical analysis to obtain a desired result.

The seminal papers of Mazar et al. (2008) and Fischbacher & Föllmi-Heusi (2013) designed self-reporting tasks (the *Matrix task* and *Die-roll task*, respectively), which became well-established paradigms for studying lying behavior and spurred a rapidly growing literature. However, these two studies, as well as many follow-up ones, exclusively investigate one-dimensional contexts, in which participants report a single outcome. Though these studies provide valuable insights regarding lying behavior (for a review see, e.g., Abeler et al., 2019), the multi-dimensionality element inherent in many real-world lying opportunities has not received attention in the literature.[1] Thus, a natural question to ask is whether (and how) a multi-dimensional context affects lying behavior. The objective of this paper is to tackle this research by investigating a two-dimensional context—which we define as a context where a person has to make two similar/identical lying decisions simultaneously—as well as contrasting it with a one-dimensional context.

To understand the effect of multi-dimensionality of a setting on lying behavior, we fundamentally assess two elements. First, we assess how people distribute lying across dimensions in a two-dimensional setting. Second, we assess whether the possibility of lying in more than one dimension affects the overall level of lying. Thus, we aim to answer two fundamental questions: (i) In a bi-dimensional setting, can we deduce lying behavior in one dimension by extrapolating lying behavior observed in the other dimension? (ii) Relative to a one-dimensional setting, does a bi-dimensional setting increase, decrease, or not affect the overall level of lying?

Our research questions emanate from research on one-dimensional lying. The central result in the growing experimental literature on lying behavior is that people lie for their own benefit, but commonly do not take full advantage of the opportunity for lying (for literature reviews and meta-studies see Rosenbaum et al., 2014; Jacobsen et al., 2017; Abeler et al., 2019; Gerlach et al., 2019). More specifically, the observed pattern in experiments consists of three types of people: (i) truth-tellers, who report observed outcomes, (ii) partial-liars, who lie, but not to the full extent, and (iii) extreme-liars, who lie to the full extent possible and thus earn the maximum payoff possible (e.g., Fischbacher & Föllmi-Heusi, 2013).

To explain this heterogeneity in lying behavior, the decision to lie has been modelled as a trade-off between the material gains from lying and the costs of telling a lie, which can be disaggregated into direct lying costs and reputational costs (Abeler et al., 2019; Gneezy et al., 2018; Khalmetski & Sliwka, 2017). Direct cost caused by violating a social or personal norm of honesty can explain individual differences in the decision to lie because lying can be too costly for some people but affordable to others (Abeler et al., 2019). Further, partial lying can be explained by reputational costs, which are higher the more obvious a lie is (Mazar et al., 2008;

[1] The recently created website http://www.preferencesfortruthtelling.com allows for visually identifying this limitation of the literature.

Fischbacher & Föllmi-Heusi, 2013; Abeler et al., 2019; Dufwenberg & Dufwenberg, 2018). Being seen by others as a liar negatively affects people's social image (when reports are observed by an audience), as well as their self-image because people can also act as their own audience even when reports are not observed by an audience (Dufwenberg & Dufwenberg, 2018).

In a multi-dimensional setting, people have the option to lie about multiple reports that jointly determine the final payoff. Thus, lying in multiple dimensions can take two forms: (i) lie partially in each dimension, and (ii) distribute lies unevenly *across dimensions*. That is, people willing to lie can combine false and truthful reports instead of telling a partial lie in each dimension. For the sake of simplicity, let us assume two dimensions and that a given aggregate self-benefit of lying can be achieved either by partially lying in both dimensions or by lying in only one dimension. If each individual lie causes direct lying costs, a person has the choice to incur two times the direct cost of a partial lie or to only incur the direct cost on one dimension and to report truthfully on the other. Accordingly, lying on only one dimension is associated with lower direct lying costs, and thus preferred, if the direct costs of this (possibly bigger) lie in one dimension is lower than twice the cost of the partial lie.[2]

Another potential motive for distributing lies unevenly across dimensions relates to image concerns. When people make multiple separate reports, image-related costs of lying will depend on people's beliefs about whether the reports, taken together, will be seen as truthful. Here, different scenarios are possible. First, people could expect reports to be evaluated jointly (Is the sum of what is reported true or a lie?). If so, dimensionality should not affect lying behavior because it does not matter how lying is distributed across dimensions, but rather how the reports are regarded together. Second, reports could be instead evaluated separately to assess whether the person making the reports is a liar (Is the first report true or a lie? Is the second report true or a lie? And so forth.).

Effron and Monin (2010) find that people are willing to forgive an (ambiguous) immoral act if the person under suspicion also performed a good deed. Thus, if a person who performs an immoral act (perceived by others as ambiguous) couples it with a moral act, the others are more willing to tolerate the possible occurrence of a transgression. In a multi-dimensional lying setting, the latter evidence would predict that other people assess a person's self-reports individually, and self-reports that appear truthful can compensate for self-reports that appear a lie. In short, according to this explanation, making a truthful report could reduce or even cancel out the reputational costs of lying in the other report.

In addition to direct and reputational lying costs, a motive to observe people distributing lies unevenly across dimensions might also be that truthful reports can act as a justification for lying in other dimensions. This conjecture is consistent with the literature on moral balancing (e.g., Blanken et al., 2015; Mullen & Monin, 2016).

[2] This holds true for fixed direct lying costs as in Khalmetski and Sliwka (2017), as well as for lying costs that take into account the size of the lie in a manner similar to Gneezy et al. (2018), as long as the marginal cost of increasing the size of a lie is decreasing.

Research in both economics and psychology has shown that people balance moral and immoral decisions (e.g., Dolan & Galizzi, 2015; Blanken et al., 2015; Mullen & Monin, 2016; Merritt et al., 2010). With regards to lying, Ploner and Regner (2013) find that people make larger donations after having cheated, and Cojoc and Stoian (2014) show that people lie more if they know that they can donate to charity later. The theory underlying the balancing of moral and immoral actions is that such balancing allows an individual to maintain a positive self-image because the moral action counteracts the negative effect that an isolated immoral action would have on his or her self-image.

To analyze the effect of dimensionality on lying, we conduct a controlled laboratory experiment with two treatments: the *One-dimensional* treatment (1DT) and the *Two-dimensional* treatment (2DT). More specifically, we implement a die-roll task with two dice. In both treatments, participants are paid according to the sum of two dice, which are rolled simultaneously. In 1DT, we ask participants to report the sum of the two dice in a single report. In 2DT, to induce bi-dimensionality, we ask participants to report each die separately. The key distinction between the two treatments is therefore the number of reports. Importantly, we preserve complete symmetry in dimensions so that we isolate the effect of dimensionality per se—i.e., the effect of making two possibly false reports.

Moreover, in light of the growing demand for validation by combining different types of data in experimental work (e.g., Gneezy & Imas, 2017), we also conduct a lab-in-the-field experiment as a validation check for the results obtained in the laboratory. Specifically, we conduct 1DT and 2DT with teenage students in their familiar school environment.[3]

Our paper provides rigorous evidence regarding the effect of multi-dimensionality on lying behavior. We find that people *distribute lies unevenly across dimensions*. Specifically, participants lie considerably more regarding the lower die. Regarding the overall level of lying, we do not find a significant difference between 1DT and 2DT. In other words, when two simultaneous lying decisions are possible, lying occurs to a greater extent in one decision than in the other, but that pattern does not lead to more lying behavior overall. In fact, it is possible that participants employ the same strategy in 1DT since reports here are an aggregate of two outcomes. If so, our findings in 2DT show the importance of considering all lying decisions by themselves. Importantly, all these results are validated in the lab-in-the-field experiment.

As far as we know, our paper is the first study in experimental economics to assess the pure effect of dimensionality on lying behavior. Shu et al. (2012) conduct an experiment in which people can lie in two different items. However, we cannot learn much from that study about how behavior is related across the two dimensions because it is not the focus of their analysis. The closest works to our paper are the parallel studies of Geraldes et al. (2021) and Barron (2019). Geraldes et al. (2021) investigate lying behavior using a two-dimensional context, but their aim is

3 To make the validation check more stringent, we not only change the type of experiment but also use a population with different demographics.

to understand how big and small lies are interconnected. This means that the two "dimensions" in their setting are not symmetrical, as the stakes component of the two lying opportunities are different. They find that people who engage in big lies also engage in small lies. Barron (2019) also offers a high-stakes and a low-stakes lying opportunities. The author finds that people over-report on the high-stakes report but under-report on the low-stakes report.

The remainder of this paper is organized as follows. In Sect. 2, we describe the experiment and formulate our hypotheses. We discuss our empirical results in Sect. 3 and conclude in Sect. 4.

2 Experimental design

2.1 Treatments

The experiment is designed with a two-fold purpose: (i) assess how participants distribute lying behavior across dimensions, and (ii) compare the overall level of lying in a two-dimensional setting (2DT) with the level of lying in a one-dimensional setting (1DT). Regarding the first purpose, to carry out the assessment as cleanly as possible, we purposefully design the two dimensions to be identical. Moreover, to serve the second purpose, we design the range of the aggregate level of lying in 2DT to be equivalent to the range of the level of lying in 1DT. The fundamental feature of our experiment is therefore to isolate as much as possible the effect of dimensionality per se on lying behavior.[4]

More specifically, we used the well-established die-roll task (Fischbacher & Föllmi-Heusi, 2013), but with two dice. In 1DT, participants were instructed to roll both dice in one hand and to report the sum of the two dice. In 2DT, participants rolled the two dice simultaneously but in separate hands and reported the outcome of each die separately. For the sake of clarity, in each treatment, the instructions included a short video to illustrate how to roll the dice. A participant's payoff was determined according to the reported eyes, where 1 eye is worth €0.50. This means that payoffs could be between €1.00 and €6.00 in both treatments.

In both treatments, participants repeated the die-roll task over ten rounds and one round was selected randomly for actual payment at the end of the experiment. The repetition element of our experiment allows us to: (i) collect more data while still assessing the first round as if it were a one-shot experiment[5], (ii) assuming lying-balancing across dimensions occurs in the first round, test whether it persists when lying-balancing over rounds is possible, and (iii) construct a measure of lying at

[4] Evidently, the absence of a substantial difference between the two dimensions makes our experiment a conservative test of the effect of dimensionality on lying behavior.

[5] The nine repetitions of the task were announced only after the first round. Therefore, the first round only allowed participants to engage in distributing lies unevenly across dimensions, while the subsequent nine rounds also allowed participants to balance lying across rounds. See Online Appendix B2 for a discussion of the pros and cons of the incentive scheme we implement in relation to the (slight) deception it generates.

the individual level. We describe the details of the latter two aspects in the Results section.

2.2 Hypotheses

Based on the literature that we discussed in the Introduction, we expect participants in 2DT to distribute lies unevenly across dimensions. We conjecture that, under the specific assumptions discussed in Sect. 1, such behavior allows participants to: (i) reduce direct lying costs, (ii) decrease the likelihood of being seen as a liar using one seemingly truthful report in one dimension to make a lie in the other dimension more credible, and (iii) "*do*" *moral balancing*—i.e., use a truthful report in one dimension as a justification for a false report in the other dimension. Therefore, our first hypothesis is:

Hypothesis 1 In 2DT, people distribute lies unevenly across dimensions.

Moreover, assuming the reputational concerns and moral balancing motives discussed in the Introduction do exist, if the magnitude of these two effects is substantial, distributing lies unevenly across dimensions could cause overall lying in 2DT to be higher than in 1DT. First, if participants believe that submitting a truthful report in one dimension sufficiently covers up a lie in the other dimension, they might be more willing to lie extremely in the dimension in which they submit a false report, and thus make, on average, higher reports than in 1DT. Second, analogous to the first argument, if making a truthful report in one dimension can sufficiently extenuate participants' moral concerns of lying, then they will consider legitimate very high reports on the other dimension. Thus, the moral balance motive is another reason to expect an increase in the average reports made in 2DT.[6] Accordingly, our second hypothesis is:

Hypothesis 2 The overall level of lying in 2DT is higher than in 1DT.

2.3 Experimental procedure

The experiment was conducted at the Experimental Laboratory for Sociology and Economics in Utrecht (The Netherlands). Participants were recruited using ORSEE (Greiner, 2015) among students of Utrecht University. In total, 139 participants took

[6] Strictly speaking, it is possible that participants in 1DT motivate their report with moral balancing, i.e., we cannot rule out that participants manipulate individual reports before calculating the sum of the two dice in 1DT. Nevertheless, even if we assume that moral balancing was employed to make the single report in 1DT, we see two reasons for the moral balancing effect to be stronger in 2DT. First, moral balancing in 2DT is more salient because it is made explicit in two actions. Second, building on the self-signalling literature (Bénabou & Tirole, 2004; Bodner & Prelec, 2003), people receive diagnostic value on their own type from actions, but not intentions. Accordingly, moral balancing in 2DT can yield a signal on their own moral character, whereas the individual report in 1DT cannot because it is a single lie.

part in the experiment.[7] The participant average age was 23.6 years, and 74% were female. Demographic variables did not differ significantly between treatments. Upon arrival at the laboratory, participants received all instructions on-screen using oTree software (Chen et al., 2016) and were randomly assigned to 1DT or 2DT, resulting in group sizes of 68 and 71 participants, respectively. Including a participation fee of €2.00, the average payment per participant was €5.85. Each session lasted around 30 min.

3 Results

We start by analyzing the level of lying for the die-roll tasks in our setting. Figure 1 illustrates the distribution of the reported sums in each treatment across the ten rounds. We observe that sums on the left side of the distribution are reported less frequently, while sums on the right side of the distribution are reported more frequently than expected under truth-telling, respectively.

Averaged over the ten rounds of 1DT, participants reported an average sum of 7.3971 and show a significant deviation from the expected outcome of rolling two fair dice ($p = 0.001$, Kolmogorov–Smirnov one sample test (KS); $p = 0.030$, two-sided Wilcoxon signed-rank test (WSR)).[8] We reach the same conclusion when pooling the 10 rounds of 2DT, in which participants reported an average sum of 7.6394 ($p < 0.001$, KS; $p < 0.001$, WSR).[9]

3.1 Lying behavior: 1DT vs. 2DT

Averaged over the ten rounds of 2DT, participants' reports are higher than in 1DT, but this difference is not significant ($p = 0.067$, Mann–Whitney (MW)).[10] Additionally, we run a regression analysis which is reported in Table 1. In Models (1), (2), and (3), the dependent variable is the average report over all rounds minus 7. The significance of the constant in Model (1) indicates significant levels of over-reporting in the pooled data. In Model (2), we run a bivariate regression including a 2DT dummy. The non-significance of this dummy ($p = 0.197$) corroborates the result that average reports do not differ significantly between treatments. This

[7] This sample size provides a power of 0.8 to detect an effect size comparable to moral licensing effects (Blanken et al., 2015), which is a Cohen's d of 0.3–0.35, and a power > 0.8 to detect a medium effect size (Cohen's d of 0.4–0.5), as suggested by Schindler and Pfattheicher (2017). Minimum detectable effect sizes of the collected data are available upon request.

[8] The distribution of two dice rolls has a bell-shaped distribution. Thus, as the KS is less sensitive to differences in the tail of distribution, we complement the test with the WSR test.

[9] The distribution of round 1 is included in the Online Appendix. Reporting in the first round is not significantly different from what is expected under truth-telling ($p > 0.1$, WSR for both tests) and does not differ significantly between treatments ($p = 0.216$, Mann–Whitney).

[10] Throughout this section, we consider a test to be statistically significant only if $p < 0.05$. Moreover, the results we claim in this section are re-tested in the replication experiment that we will report in Sect. 3.4.

finding is robust when including demographic control variables (in Model (3), the *p* value associated with the 2DT coefficient is 0.064). Collectively, these results do not support Hypothesis 2.

Moreover, we also assess lying at the individual level based on the repetition feature of our experiment. Despite not having a hypothesis regarding the fraction of liars in each treatment, we consider it important to explore this aspect of our design. To this end, we construct a dummy variable that indicates a *likely liar*. More specifically, this variable indicates whether the average report of a participant has a chance of 10% or lower of occurring under truthful reporting—i.e., whether the average report is higher than the 90%-quantile of the truthful distribution of two die rolls over ten rounds. In 1DT, 22% of participants fall into this category, compared to 30% in 2DT. The difference between treatments is not significant (Chi^2 test (CT), $p = 0.312$).[11]

Result 1 (i) Reported sums in 2DT are not significantly different than reported sums in 1DT. In other words, we find no support for Hypothesis 2. (ii) We find that the fraction of likely liars in each treatment is not significantly different, either.

3.2 Lying behavior across dimensions in 2DT

To test lying-balancing strategies, we perform two separate analyses. First, we analyze lying behavior across the two dimensions in the first round of 2DT, as participants learned about the subsequent nine rounds only after completing Round 1. Second, we consider lying behavior across the subsequent nine rounds of 2DT so that we test lying-balancing not only across dimensions but also across rounds. We test potential lying-balancing across rounds in Sect. 3.3.

In 2DT, participants were instructed to roll one die in the left hand and the other one in the right hand. Reports on the left and right dice in the first round of 2DT are not significantly different (3.7183 vs. 3.6338, $p = 0.799$, WSR). The same holds for the subsequent nine rounds (3.7559 vs. 3.9155, $p = 0.092$, WSR). This result is not surprising. Since the two dimensions are identical, participants had no reason to prefer lying about the left die over lying about the right one, or vice versa.[12] However, this result does not necessarily imply that lying-balancing did not occur. That is, balancing across dimensions could have occurred by sometimes lying about the left die outcome and at other times lying about the right die outcome, depending on the actual outcome of each die.

[11] Both these shares are significantly different from 10% (two-sided proportion test, $p < 0.01$ for both tests. The comparison between treatments is also not significant when we compare truly likely liars only: 12% vs. 20%.

[12] An analysis of *likely liars* analogous to the one reported in Sect. 3.1 corroborates this result. Here, we constructed a dummy variable that indicates whether a participant's average report on the right (left) die has a 10% chance or lower of occurring under truthful reporting—i.e., whether the average report is higher than the 90%-quantile of the truthful distribution of one die rolls over ten rounds. On the right die, 35% of participants are classified as *likely liars*, compared to 24% on the left die. The difference is not significant (McNemar test on paired proportions, $p = 0.131$).

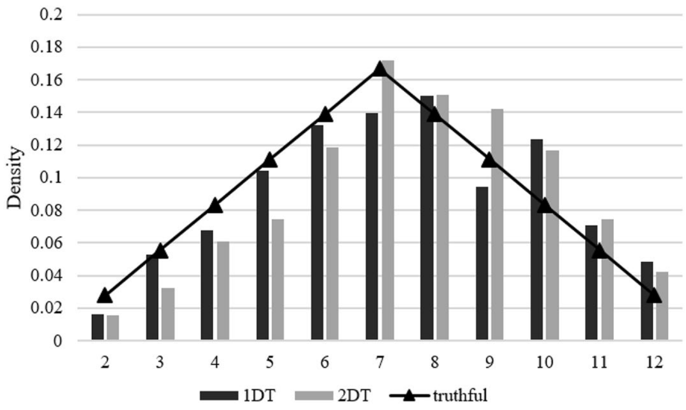

Fig. 1 Distribution of reported sums of rolling two fair dice (truthful distribution depicted as a reference)

Table 1 Linear regression models of the level of lying

Outcome variable: (average report − 7)			
	(1)	(2)	(3)
2DT		0.2424	0.3405
		(0.1870)	(0.1824)
Constant	0.5209***	0.3971**	0.4615
	(0.0937)	(0.1336)	(0.5126)
Controls	No	No	Yes
N	139	139	137[a]
R2	0.0000	0.0121	0.0401

Standard errors in parentheses.

Controls include gender, age, and studying economics.

[a]Two participants indicated gender = other.

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

To gain more insight, we sort the two reports in 2DT into lower and higher reports to test whether balancing across the two dimensions occurred based on the observed outcomes. Importantly, in the following analysis, we assume that a participant's higher report was based on the higher observed outcome and the lower report based on the lower observed outcome.[13] Under truth-telling, the expected value for the lower and higher of two dice rolls is 2.5278 and 4.4722 respectively.[14] Next,

[13] Evidently, we do not actually know whether this is the case because the die rolls are private information. It is also possible that a participant's higher report was based on the lower observed outcome. That would imply that our follow-up analysis—which finds significantly more lying about the lower die—might, if anything, be underestimating the positive gap in lying about the lower die relative to the higher die because it underestimates lying about the lower die and overestimates lying about the higher die.

[14] Equal reports on both dice are included in the analysis, with the same value being assigned to the higher and lower die, and the expected values under truth-telling are derived the same way. If we exclude ties, the results are qualitatively the same.

we analyze deviations from the respective expected value for both the lower and the higher die so that we can compare over-reporting between the two outcomes. For this purpose, we construct a variable measuring (*Report* − *expected value*) for every participant and round.

In the first round, for the ordered reports—which we designate the *higher die* and the *lower die*—participants lied significantly more regarding the outcome of the lower die. Specifically, the average deviation from the expected value under truthful reporting is higher for the lower report than for the higher report, and this difference is significant (0.3173 vs. 0.0348, $p = 0.039$, two-sided t test (TT)).[15] Regarding the subsequent nine rounds, we observe the same behavioral pattern. Participants' average reports on the lower die show significantly higher deviation from the expected value than the average reports on the higher die (0.4128 vs. 0.2586, $p = 0.017$, TT).

Arguably, the higher levels of over-reporting on the lower die could be merely an artefact of less lying being possible for the higher die. To exclude that this is driving differences in means, we also analyze the distribution of the high and low reports. Even if participants could not report values greater than six on the high die, if participants were lying to the same extent on both dice, we should find a similar shift of probability mass away from the distribution expected under truth-telling. Figures 2a and 2b show the distribution of high and low reports in Round 1.[16] The corresponding distributions for the remaining nine rounds show the same pattern (see Figure A1.2 in Online Appendix A1).[17] The distributions show the greatest deviation from the truthful distribution for under-reporting of 1 on the low die, i.e., lying on the low die took the form of adjusting very low outcomes upwards. For the high die, we find no over-reporting of 6 in the first round and more over-reporting of 5 than of 6 in the nine later rounds. In fact, the share of 6 being reported in the nine later rounds is not significantly different from the share expected under truthful reporting (two-sided binomial test (BT), $p = 0.078$). Thus, upon visual inspection, participants do not take full advantage on over-reporting 6 on the high die but that more commonly they adjust an observed die roll of 1 on the low die upwards. In a next step, we will quantify the difference in deviations from the truthful distribution.

We compare distributions of the high and low die by considering how far reports (r) differ from the truthful distribution (t). To this end, we use the Euclidean distance from the distribution and bootstrapped confidence intervals with 10,000

[15] Since both reports are centered on the expected mean, we use a t test to compare the means of the deviations.

[16] Both are not significantly different from the respective truthful distribution ($p = 0.163$ and $p = 0.999$, respectively, KS).

[17] Distributions in Figure A1.2 are significantly different from expected distributions under truth-telling ($p < 0.001$ for both, KS). To account for participant effects, we also bootstrap p values for Kolmogorov–Smirnov tests between the average high (low) die in nine rounds between the observed data and simulated data that assumes truth-telling and find the result to be robust ($p < 0.001$ for both).

replications.[18] Regarding the lower reports, the Euclidean distance to the truthful distribution is $d(r,t)_{\text{low},1} = 0.1452$ (confidence interval 0.0466–0.2438) in the first round and $d(r,t)_{\text{low},2-10} = 0.1316$ (confidence interval 0.0995–0.1637) in the subsequent nine rounds. Regarding the higher reports, the analogous figures are $d(r,t)_{\text{high},1} = 0.0785$ (confidence interval − 0.0050 to 0.1620) and $d(r,t)_{\text{high},2-10} = 0.0847$ (confidence interval 0.0541–0.1152). Thus, for the first round, we cannot reject the possibility that participants reported the higher outcome truthfully. Notably, for the nine later rounds, estimates for the low and the high dice are not included in each others' confidence intervals, and the proportional overlap is only 0.50 (Cumming, 2009). In a nutshell, the Euclidean distance analysis indicates that the distribution of reports on the low die deviates more from the truthful distribution than the distribution of reports on the high die. Thus, we conclude that the average deviations from the expected report on the high die were less than on the lower die not (only) because participants had less room to lie upwards regarding the higher die.

Finally, we again construct a dummy variable that indicates whether a participant is a *likely liar*. Here, we compute this variable for the low and for the high die, respectively. For the high die, the variable captures whether a participant's average report on the high die in the subsequent nine rounds has a 10% chance or lower of occurring under truthful reporting—i.e. whether the average report is higher than the 90%-quantile of the truthful distribution of the higher of two die rolls over nine rounds. For the lower die, we compute the analogous variable. Regarding the high die, 30% of participants made such reports, compared to 41% on the low die. The difference between the two is not significant (McNemar test on paired proportions (MNT), $p = 0.088$).[19] Overall, 49% of participants are classified to be neither a *likely liar* on the low nor on the high die, 20% are classified as *likely liars* on both dice, 21% are classified as *likely liars* only on the low die, and only 10% are classified as *likely liars* only on the high die. In short, the analysis of *likely liars* reveals that more lying in the low die is driven by the intensive margin (i.e., liars lie more about the low die), since the fraction of likely liars is higher for the low die, but the difference is not significant.

In short, our results support Hypothesis 1.

Result 2 Participants distribute lies unevenly across dimensions, with more lying in the reported lower outcome.

[18] Approaches of earlier literature to compare lying in different tasks are not applicable to this situation. The standardization method of Abeler et al. (2019) assumes symmetric distributions under truth-telling, and the lying calculator of Garbarino et al. (2016) is limited to binary outcomes.

[19] Again, the fact that this result is marginally significant with the laboratory sample demanded further scrutiny. The replication experiment with the school sample clearly supports that the fraction of likely liars in the lower and higher reports are not significantly different ($p = 0.285$, MNT).

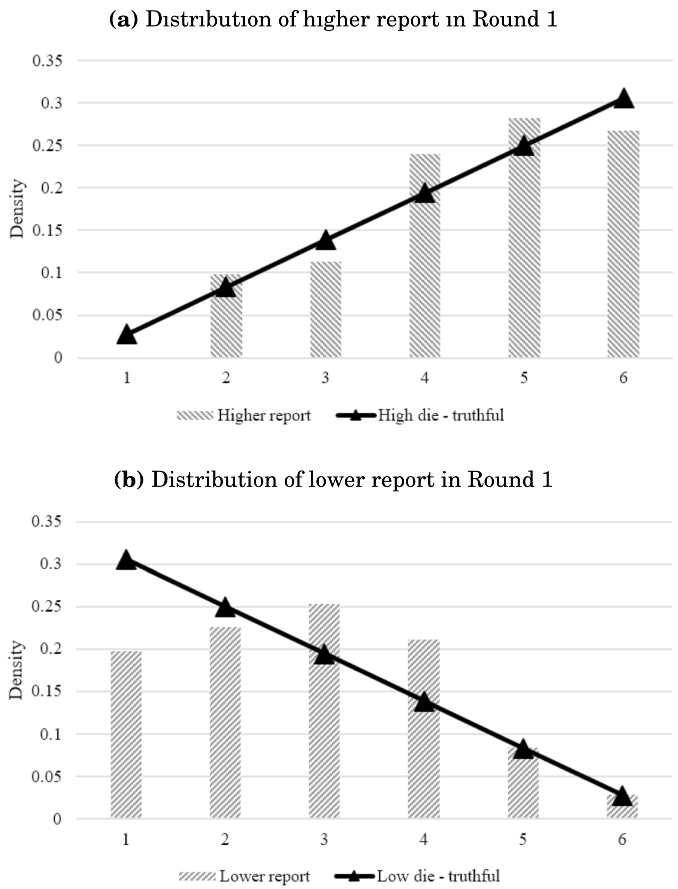

Fig. 2 Distribution of lower and higher reports in Round 1 of 2DT (truthful distribution depicted as a reference)

3.3 Behavior over rounds

In Table 2[20], we see that the round effect is insignificant in both treatments, which indicates that participants did not change their behavior over rounds in a systematic fashion. Also, there is no evidence of participants balancing lying across rounds in the form of alternating a lie with a truthful report, which would have been the case if we had a significant negative coefficient of the lagged outcome variable.[21]

[20] More information on the model specifications are included in Online Appendix A1.

[21] For persistent liars, lying in earlier rounds is positively associated with lying in later rounds. However, since the Arellano–Bond estimator eliminates individual effects, we do not observe a positive coefficient for the lagged outcome variable.

Table 2 Two-step system GMM with lagged levels ($t-2$, $t-3$ and $t-4$) of the dependent variable as instruments

Outcome variable:	1DT diesum		2DT diesum		2DT lowdie		2DT highdie	
l. Outcome	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	− 0.028	− 0.0218	0.0659	0.0527	− 0.0284	− 0.0752	0.1287*	0.1134*
	(0.0611)	(0.0569)	(0.0521)	(0.0525)	(0.0518)	(0.0501)	(0.0579)	(0.0568)
Round		− 0.042		− 0.0329		− 0.0036		− 0.01
		(0.0360)		(0.0364)		(0.0241)		(0.0227)
Constant	7.6168***	7.8212***	7.2170***	7.5132***	2.9902***	3.1680***	4.1773***	4.3047***
	(0.4855)	(0.4609)	(0.4181)	(0.4582)	(0.1821)	(0.2304)	(0.2711)	(0.2993)
Arellano–Bond serial	− 0.24	− 0.18	0.19	0.1	0.3	− 0.14	− 0.1	− 0.18
correlation test [*p* value]	[0.813]	[0.856]	[0.846]	[0.917]	[0.762]	[0.891]	[0.923]	[0.861]
Instruments	30	31	30	31	30	31	30	31
Sample size	612	612	639	639	639	639	639	639

Robust standard errors using Windmeijer correction in parentheses.

*$p < 0.05$, **$p < 0.01$, *** $p < 0.001$

Result 3 There is no evidence of uneven distribution of lies across rounds.

3.4 Validation check: lab-in-the-field experiment

To test the validity of the results, we replicated the lab experiment in the field with a different population. Specifically, we conducted 1DT and 2DT with adolescents in a German school using the same die-roll tasks as in the lab experiment. We chose a high school to implement the lab-in-the-field experiment because it offered the two features that we aimed for conducting a stringent validation test: (i) a field setting, and (ii) a different population.

In a nutshell, the results in the school population are completely in line with our findings in the lab. First, we again find no significant difference in the level of lying between 1DT and 2DT. While the 2DT regressor has a marginally significant positive effect in Model (3) of Table 1, the replication shows a non-significant 2DT regressor, and its directional effect is even negative (see Table A2.1 in Online Appendix A2. The p-value associated with the 2DT coefficient is 0.620 in Model (2) and 0.607 in Model (3)). Second, lying about the right and left dice does not differ significantly, but again, we observe more lying in the lower report. Finally, we find no evidence of round effects or uneven distribution of lies across rounds.[22]

Result 4 The results of the lab experiment are validated in a different population.

4 Discussion and concluding remarks

This paper contributes to a growing body of literature on lying behavior. To the best of our knowledge, the present research is the first attempt in experimental economics to isolate the effect of splitting reporting into two on individual lying behavior.

First and foremost, our results uncover an interesting lying trait in the two-dimensional context: People distribute lies unevenly across dimensions. More specifically, people systematically lie more in the dimension in which there is more to gain. Importantly, this laboratory result is replicated in a lab-in-the-field experiment with participants possessing different characteristics. Additionally, we find no significant difference in the levels of lying in 2DT compared to 1DT.

Our results can be linked to the mechanisms that we discussed in the Introduction. First, people lying more in the dimension with higher possible gains can be explained by people having intrinsic lying costs. Second, our results also accommodate a reputational costs explanation, based on the combination of the likely truthfulness of the individual reports. That is, a mechanism in which one report that seems truthful

[22] The detailed experimental procedure and results of the lab-in-the-field experiment are presented in Online Appendix A2.

by itself is used to offset the reputational damage of a high second report could explain the pattern of our results. Third, our results also fit in with the occurrence of moral balancing. If a truthful report in one dimension is needed to balance a lie in the other report, it is payoff-maximizing to lie in the dimension with higher possible gains. Future research is needed to disentangle the relevance of these mechanisms.

An important implication of our results is that lying behavior does not necessarily extrapolate from one dimension to another because people distribute lies unevenly across dimensions. To illustrate the implications of our results, consider the example of tax declarations. First, to detect fraud in tax reports, authorities should focus on the item(s) for which the self-benefits of misreporting are higher. Secondly, if fraud is detected in a specific item, authorities should not necessarily assume that the same pattern holds regarding other items.

While this paper is an important first step in understanding multi-dimensional lying, our research yields questions in need of further investigation. First, an obvious follow-up is to test whether the uneven distribution of lies across dimensions holds in a setting with several dimensions. Second, we did not find support for the hypothesis that a two-dimensional context increases overall lying behavior. However, a possible explanation for the observed similar level of lying in both contexts could be that the contexts were too identical. In particular, we cannot rule out that participants in 1DT also used an uneven distribution of lies strategy since they also rolled two dice. Nevertheless, if the latter conjecture is true, that would be a further argument for the importance of studying a multi-dimensional reporting context: It uncovers traits of lying behavior that are concealed in a one-dimensional reporting context.[23] Third, from a laboratory experimental methodological perspective, it was advisable as a first step to conduct a conservative test of the effect of dimensionality on lying behavior. A natural and relevant extension to our work is to study lying in a multi-dimensional setting where the dimensions are dissimilar.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40881-022-00121-3.

Acknowledgements The authors thank the participants of the EDS Seminar (Utrecht University), ESA World Meeting, TIBER Symposium (Tilburg University), Nordic Conference on Behavioural and Experimental Economics, M-BEES (Maastricht University), and the ZEW/University Mannheim Experimental Seminar for their invaluable comments. The authors are particularly indebted for the comments of Kai Barron, Maria Bigoni, Dirk Engelmann, Jan Potters, and two anonymous reviewers. The authors gratefully acknowledge financial support from Utrecht School of Economics.The replication material for the study is available at https://osf.io/c298e/?view_only=b3273d39295f47a99a14d503c7cab3c2.

Funding Utrecht School of Economics.

Availability of data and material All data and materials available upon publication.

Declarations

Conflicts of interest No conflicts of interest.

[23] An alternative explanation for not observing a higher level of lying in 2DT is that a costly moral action in one dimension does not justify an immoral action in the other dimension (Gneezy et al., 2012).

References

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

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58*(2), 277–297.

Barron, K. (2019). Lying to appear honest. Discussion Paper SP II 2019–307, WZB.

Bénabou, R., & Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy, 112*(4), 848–886.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin, 41*(4), 540–558.

Bodner, R., & Prelec, D. (2003). The diagnostic value of actions in a self-signaling model. *The Psychology of Economic Decisions, 1,* 105–123.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9,* 88–97.

Cojoc, D., & Stoian, A. (2014). Dishonesty and charitable behavior. *Experimental Economics, 17*(4), 717–732.

Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine, 28*(2), 205–220.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology, 47*(Supplement C), 1–16.

Dufwenberg, M., & Dufwenberg, M. (2018). Lies in disguise—A theoretical analysis of cheating. *Journal of Behavioral and Experimental Finance, 175,* 248–264.

Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions? *Personality and Social Psychology Bulletin, 36*(12), 1618–1634.

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

Garbarino, E., Slonim, R., & Villeval, M. C. (2016). Loss aversion and lying behavior: Theory, estimation and empirical evidence. GATE Working Paper 1631.

Geraldes, D., Heinicke, F., & Kim, D. G. (2021). Big and small lies. *Journal of Behavioral and Experimental Economics, 91,* 101666.

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin, 145*(1), 1.

Gneezy, U., & Imas, A. (2017). Lab in the field: Measuring preferences in the wild. In: *Handbook of economic field experiments* (Vol. 1, pp. 439–464). Elsevier, Amsterdam.

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behavior. *Management Science, 58*(1), 179–187.

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

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114–125.

Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2017). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys, 32*(2), 357–387.

Khalmetski, K., & Sliwka, D. (2017). Disguising lies—Image concerns and partial lying in cheating games. CESifo Working Paper Series 6347.

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.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4*(5), 344–357.

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology, 67*(1), 363–385.

Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization, 93*(Supplement C), 374–383.

Roodman, D. (2009). How to do Xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal, 9*(1), 86–136.

Rosenbaum, S. M., Billinger, S., & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology, 45,* 181–196.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology, 69,* 172–177.
Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*(38), 15197–15200.
Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics, 126*(1), 25–51.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.