Noname manuscript No.
(will be inserted by the editor)

Lies, Labor, and Luck

Comparing Lying in Real-Effort and Luck Tasks

Simon D. Halliday · Oumayma Koulouh ·

Received: date / Accepted: date

Abstract We study the differences in lying behavior between real-effort and luck-based tasks. While many papers use luck-based tasks to study deception, recent research shows that individuals may behave differently when the payoffs result from luck or from real effort. We conduct an experiment (n = 114) with a 2 by 2 factorial design in which we observe lying behavior at the individual level. We compare lying in luck-based and real-effort tasks and find that the proportion of people is constant across the tasks. We also compare two real-effort tasks, one of which contains a greater luck component and find no differences across the two tasks. **JEL Classification:** C91, D01, D82, H26

Keywords Deception · Lying · Luck · Real Effort ·

Simon D. Halliday

Department of Economics, Smith College, Wright Hall 234, 5 College Lane, Northampton, MA, 01062, USA

E-mail: `shalliday@smith.edu`

Oumayma Koulouh

Deutsche Bank, Corporate Finance, 60 Wall Street, New York, NY, 10005

E-mail: `oumaymakoulouh@gmail.com`

1 Introduction

Many economic activities involve sharing private information in contexts where the stakes are significant. Whether it's an employee reporting the hours worked on a project, a real estate broker sharing their best advice with clients, or an individual filling out their insurance claims, people sometimes find it in their best interest to lie about private information they observe. For example, Mazar and Ariely (2006) argue that deception leads to a loss of hundreds of millions of dollars for the US economy every year, through loss of tax revenues and wages for example. Understanding the drivers of dishonesty therefore remains an important topic of study.

Recent work suggests that a person's choice to behave deceptively is sensitive to the nature of the task (Kajackaite 2018). The bulk of the economics experiments on deception focuses on luck-based tasks, such as a die roll or a coin toss, in which the outcome depends solely on luck. If participants behave differently depending on the source of income, then behavior in luck-task experiments might not characterize behavior in real life because most tasks in real-life contain a real-effort component. In the context of deception, we expect to have a higher proportion (extensive margin) and magnitude of lies (intensive margin) in luck-based tasks than in real-effort tasks. This is due primarly to the higher mutability of luck tasks (Kahneman and Miller 1986).[1]

We study the effect of task type on deceptive behavior comparing luck tasks and real effort tasks. We use two real-effort tasks: a mathematically-based matrix task and a spelling-error detection task inspired by existing experimental designs (Mazar, Amir, and Ariely 2008; Ariely and Wertenbroch 2002). In the United States, many students perceive mathematical ability to be a result of innate ability rather than improvement through effort (Uttal 1997; Devlin 2000), whereas effort is considered more important than ability in achieving good spelling (Rankin, Bruning, and Timme 1994). Subjects may therefore perceive the matrix task as relatively more luck-based than the spelling error task.

Our contribution is two-fold. First, we extend the work of Kajackaite (2018) who finds that deceptive behavior in a luck-based task is different than behavior in a real-effort task. Instead of inferring lying from results of a control treatment, we detect lying at the individual level using a method adapted from Chao and Larkin (2017). Second, we compare lying in real-effort and luck tasks, as well as differences between two different real-effort tasks. It is important to investigate whether lying behavior changes within real-effort tasks depending on whether they include a perceived luck component independent of choices in purely luck-based tasks.

Each luck task had a distribution of numbers mirroring that of the number of true outcomes in the corresponding real-effort task. For the luck tasks,

[1] Researchers argue that it is much easier to imagine a different outcome when the outcome is determined by luck compared to when it's determined by real-effort. Various papers have shown that higher mutability leads to more lying (Shalvi, Eldar, and Bereby-Meyer 2012; Shalvi et al. 2015).

participants were asked to select one piece of paper from a stack of papers and report the number they picked. Participants get paid a dollar for the value of the number they report in $[1, 20]$, which is a formulation that Kajackaite (2018) uses to correspond roughly to the format of Fischbacher and Föllmi-Heusi (2013).

Our experiment data shows statistically significant incidence of lying in the spelling error task, and no statistically significant incidence of lying in the other three tasks. Comparing lying in real-effort and luck tasks, the average lies are at least twice as large in the luck tasks. Unlike the results from Kajackaite (2018), this difference is not statistically significant. We also find no significant difference between the two real-effort tasks.

While research focused on the impact of non-task related variables, little research has compared how the propensity to lie changes across different tasks. To our knowledge, only Kajackaite (2018) compares lying in luck and real-effort tasks. Kajackaite (2018) detects lying by comparing the distribution of reported outcomes to outcomes from a control group. Our design, which is based on Chao and Larkin (2017), allows us to detect lying at the individual level and therefore to identify both the intensive and extensive margins more precisely. We also extend the work on luck versus real-effort tasks by comparing differences in lying behavior within two real-effort tasks, one of which may be perceived to include a greater luck component.

While our experiments do not provide information about the subjective lying cost parameters in an individual's utility function, the experiments allow us to draw some conclusion on how the parameters would change depending on the nature of the task. The two hypotheses we are testing are the following: first, people incur a higher cost of lying in real-effort tasks compared luck effort tasks; second, people incur a higher cost of lying in real-effort tasks that contain a higher perceived luck component. Therefore, in the context of our experiment, we expect more lying to occur in the matrix test.

2 Experimental Design

Participants were asked to complete one of four tasks- two real-effort tasks and two based on luck. We call these tasks Real Effort I (matrix task), Real Effort II (spelling task), Luck 1 (distribution based on Real Effort 1), and Luck II (distribution based on Real Effort II). For Real Effort I and Real Effort II, subjects were given 5 minutes for the real-effort task. The first real-effort task (Real Effort I) is based on Mazar, Amir, and Ariely (2008) where subjects solved mathematical matrices. Real Effort I represents the real-effort task with a higher perceived luck component as it involves a mathematics-related task (Uttal 1997). The second real-effort task (Real Effort II) involved detecting spelling errors, inspired by Ariely and Wertenbroch (2002). Examples of both tasks are included in the appendix. The spelling task used text generated from a post-modern text generator so that comprehension was irrelevant to spelling error detection.

R1	Number of Subjects	25
	% Female	48%
	Average age	19.80
	Average pay	10.68
	Average lie	0.24
L1	Number of Subjects	30
	% Female	59%
	Average age	19.87
	Average pay	12.93
	Average lie	0.47
R2	Number of Subjects	29
	% Female	47%
	Average age	20.10
	Average pay	13.34
	Average lie	0.31
L2	Number of Subjects	30
	% Female	70%
	Average age	19.69
	Average pay	15.20
	Average lie	0.83

Table 1 Summary of experimental conditions

The third task (Luck I) is a luck-based task, in which participants were given a stack of stapled papers with numbers between 1 and 20. The distribution of numbers mirrored that of the true number of solved matrices in Real Effort I. Participants were asked to select one piece of paper and remember the number it contained. The fourth task (Luck II) is similar to Luck I, but the distribution of numbers mirrored that of the number of true spelling errors in Real Effort II. All experiment materials are included in the Appendix.

Table 1 summarizes the experimental conditions of each treatment. One session was run for each treatment.

To detect lying, we use the methodology from Chao and Larkin (2017). The [Blinded for Submission] Experimental Economics Laboratory contains blocks of three seats so we numbered each adjacent seat and gave each block of seats a shared bin in which they disposed of all materials by the end of the experiment. For the real-effort tasks, we gave the three participants different colored pens. For the luck tasks, the stacks of paper were printed in three different colors. The contents of the bin allowed us to detect lying at the individual level at the end of the experiment by connecting the pen colors to the seat number. A participant's identity could not be connected to their seat number, so their behavior remained confidential.

3 Results

3.1 Description of the sample

For each treatment, we collect data about the participant's demographic data, true outcome, reported outcome, and relevant treatment. We define two variables to track lying behavior. The first is a binary variable measuring the extensive margin – whether participants lied or not. The second measures the intensive margin, that is, we measure the size of the lie by the difference between the reported outcome and the true outcome. Further, we classify lies into 2 categories. A subject *partially* lies when the reported outcome is greater than the true outcome but the reported outcome is less than 20, the maximum possible outcome. A subject *full extent* lies when the reported outcome is greater than the true outcome and the reported outcome is equal to 20. For each treatment, we compute the magnitude of lies as the difference between the reported outcome and the true outcome.

In the discussion below, we show support for the the following main results:

1. The only task in which there is statistically significant incidence of lying (on the extensive margin) is Luck II.
2. The incidence of lying is not statistically significantly different compared across treatments.

Figure 1 shows the average reported outcome by task and treatment. Figure 2 presents the average lie by treatment (intensive margin). Figure 3 shows the distribution of real and reported outcomes by treatment.

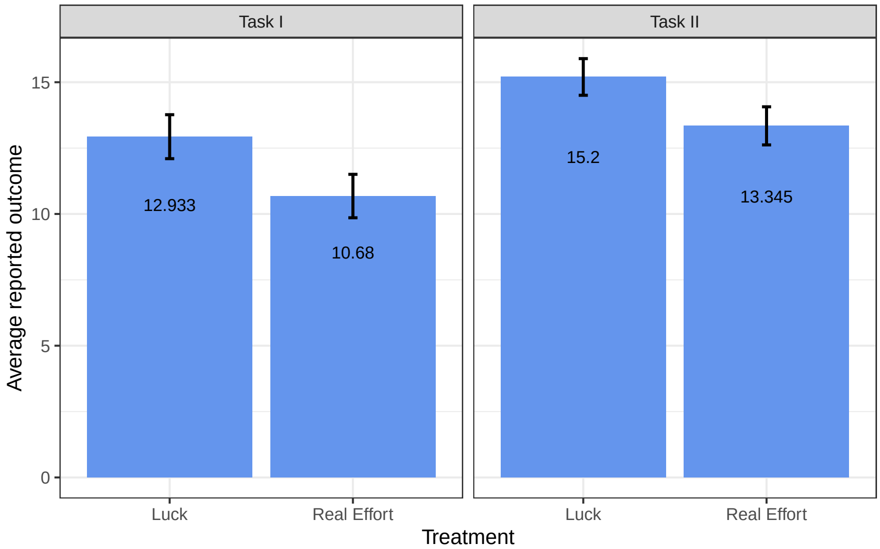

Figure 1. Average reported outcome by treatment and task

Note: Error bars are standard errors of the mean.

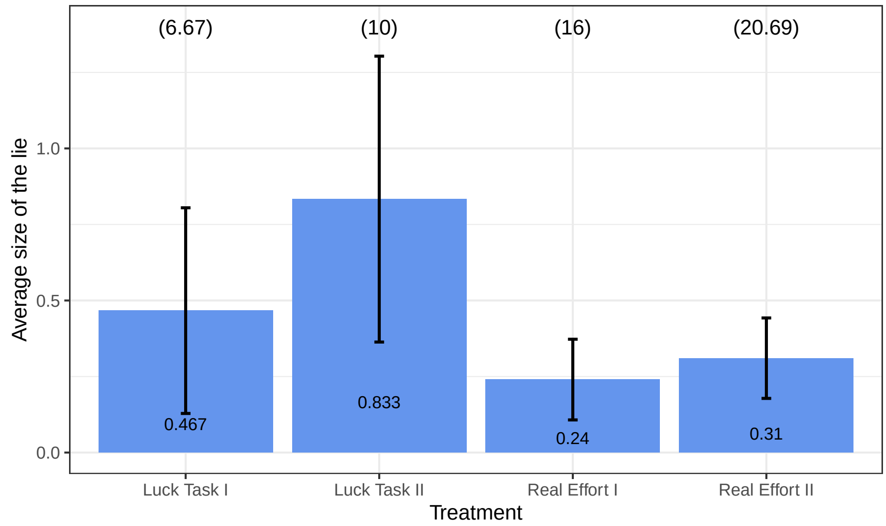

Figure 2. Average lie per treatment

Note: Error bars are standard errors of the mean.
Numbers in parentheses are incidence of lying as a proportion.

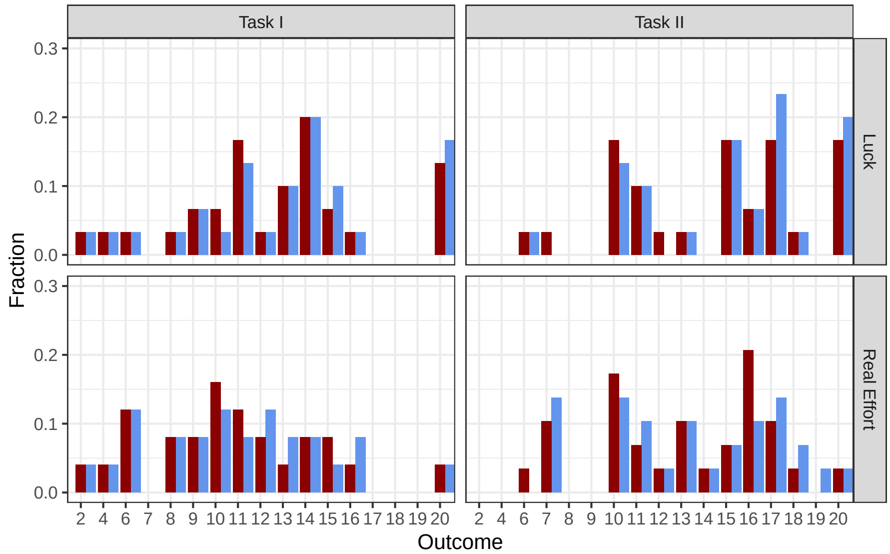

Figure 3. Distribution of actual and reported outcome by treatment

Note: The blue bars represent the reported outcomes and the red bars represent the true outcomes.

3.2 Lying across the four treatments

In the Real Effort I (RI) task, participants reported solving 10.68 matrices on average (SD=4.12). The average number of true solved matrices is 10.48 (SD=4.03). Testing for the presence of lying using the extensive margin binary variable, we find that lying is not significant (p=0.11, Fisher's exact test), with

4 participants (16%) lying. All lying in RI was partial lying with magnitudes of 1 or 3.

In the Real Effort II (RII) task, participants reported finding 13.34 errors on average (SD=3.98). The average number of true detected errors is 13.07[2] (SD=3.76). Testing for the presence of lying on the extensive margin, we find that there is statistically significant lying (p=0.023, Fisher's exact test), with 6 participants (20.7%) lying. All lying in RII was partial lying with magnitudes varying between 1 and 3.

In the Luck I (LI) task, participants reported an average of 12.93 (SD=4.57). The true average number is 12.47 (SD=4.38). Testing for the presence of lying on the extensive margin, we find that lying is not statistically significant (p=0.49, Fisher's exact test), with 2 participants (6.7%) lying. One participant reported a partial lie of size 5, and the other reported a full extent lie of size 9 (i.e. 20 - 11 = 9).

In the Luck II (LII) task, participants reported an average of 15.20 (SD=3.81). The true average observed number is 14.37 (SD=4.05). Testing for the presence of lying using the extensive margin binary variable, we find that lying isn't significant (p=0.24, Fisher's exact test), with 3 participants (10%) lying. 2 participants reported partial lies of sizes 7 and 10, and one reported a full extent lie of size 8.

3.3 Comparing lying in luck and real-effort treatments

Comparing luck and real-effort outcomes in Task I, the average lie in the luck treatment (0.47) is almost twice the average lie in the real-effort treatment (0.24). The difference is not statistically significant (p=0.34, Wilcoxon signed-rank test).

Similarly for Task II, the average lie in the luck treatment (0.83) is more than twice the average lie in the real-effort treatment (0.31). This difference is not statistically significant (p=0.37, Wilcoxon signed-rank test) due to the high standard deviations of the average lies of the luck and real-effort treatments (SD=2.57 and SD=0.71 respectively).

3.4 Comparing lying across the two real-effort tasks

As discussed above, 16% and 20.7% of participants lied in the Real effort I and Real effort II treatments respectively, with average lies of 0.24 and 0.31 respectively. Both the difference in the extensive margin (p=0.74, Fisher's exact test) and the difference in the intensive margin (p=1, Wilcoxon signed-rank test) are not statistically significant.

[2] The average number of solved matrices is lower than the average detected spelling mistakes unlike the pilot experiments in which the two averages were similar.

3.5 Comparing lying across the two luck tasks

Comparing the distribution of the numbers provided in the stack of stapled papers in Luck I and Luck II, we find that we have a higher distribution of numbers in LII compared to LI (p=0.008). However, the level of lying isn't statistically different in both the extensive margin (p=0.65, Fisher's exact test) and the intensive margin (p=0.66, Wilcoxon signed-rank test).
This is consistent with the literature that finds that an increase in payoffs doesn't affect lying behavior (Abeler, Nosenzo, and Raymond 2016; Fischbacher and Föllmi-Heusi 2013; Mazar, Amir, and Ariely 2008).

3.6 Regression Results

We run four regressions to analyze further the extensive and intensive margins of lying. We report the results in Table 2. The explanatory variables are the same across the four models: the participant's age and the treatment in which they participated.[3]

The first two regressions estimate the probability of lying through an ordinary least squares (OLS, column 1) and probit model (coefficients in column 2 and marginal effects in column 3). Both regressions show that an increase in age decreases the probability of lying, and that the probability of lying is higher in the RII task, followed by RI, LII then LI. None of the treatment dummies in the two regressions are statistically significant, consistent with our analysis in the previous section.

The second two regressions estimate the size of lying through an OLS (column 4) and tobit model (column 5). The tobit model allows us to cater for left-censoring and right-censoring of the outcome variable at 0 and 20 respectively. The predictions over the lie over the four treatment differ across the two regressions with larger effects suggested by the Tobit regressions. The OLS model predicts that the size of the lie is the highest in LII, followed by LI, RII and RI while the Tobit model predicts that the size of the lie is the highest in RII, followed by LII, RI and LI. None of the coefficients in the two regressions are statistically significant, consistent with our analysis in the previous section.

4 Discussion and Conclusion

Overall, our experiment data shows significant lying in the real-effort spelling task, and no lying across the other 3 experimental tasks. Our results are close to those of Abeler, Becker, and Falk (2012) who found no evidence of lying in their luck experiment.

[3] We exclude gender from the regressions. Please see the appendix for results that include a dummy variable for male subjects as compared to female and non-binary subjects. The coefficient on the male dummy variable is not statistically significant.

Table 2 Regression estimates

	Extensive Margin			Intensive Margin	
	OLS	Probit	Probit MFX	OLS	Tobit
Age	−0.01	−0.07	−0.01	−0.09	−0.57
	(0.02)	(0.10)	(0.02)	(0.10)	(0.70)
Treatment LII	0.03	0.21	0.04	0.35	1.93
	(0.09)	(0.47)	(0.11)	(0.44)	(3.20)
Treatment RI	0.09	0.51	0.12	−0.23	1.96
	(0.09)	(0.47)	(0.12)	(0.46)	(3.32)
Treatment RII	0.14	0.72	0.18	−0.13	3.37
	(0.09)	(0.44)	(0.12)	(0.44)	(3.22)
Constant	0.33	−0.09		2.29	1.23
	(0.41)	(2.03)		(2.05)	(13.82)
R^2	0.03			0.03	
Adj. R^2	-0.01			-0.01	
Log Likelihood		-42.61	-42.61		-77.62

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The relatively low amount of lying makes it hard to detect differences over the treatments. While the average lies support the results from Kajackaite (2018), we find no significant differences between lying in real-effort and luck tasks. This emphasizes the dependence of lying preferences on the specific context of the experiments, even when the same experimental task is used. The difference between the results in Kajackaite (2018) and ours suggests a need for further research into comparing real-effort and luck tasks in different environments and experimental conditions. We find no significant difference between the two real-effort tasks, which contradicts our original hypothesis. Therefore, if there is more lying in real-effort tasks that contain a luck component as we had hypothesized, the effects are small enough to be overlooked in modeling behavior and designing policy.

Our experiments therefore show low incidence of lying, suggesting that people face an intrinsic cost of lying in low-stakes circumstances. Comparing lying behavior between real-effort and luck tasks, we find that the proportion of individuals choosing to lie remains constant. We also extend the comparison by using two real-effort tasks, one of which includes a higher luck component. Our results show no difference within real-effort tasks.

5 References

Abeler, Johannes, Anke Becker, and Armin Falk. 2012. “Truth-Telling: A Representative Assessment.”

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2016. “Preferences for Truth-Telling.”

Ariely, Dan, and Klaus Wertenbroch. 2002. “Procrastination, Deadlines, and Performance: Self-Control by Precommitment.” *Psychological Science* 13 (3). SAGE Publications Sage CA: Los Angeles, CA: 219–24.

Chao, Matthew, and Ian Larkin. 2017. “Why Do Goal-Based Incentives Cause Cheating? Unpacking the Confounding Effects of Goals, Social Comparisons and Pay.”

Devlin, Keith J. 2000. *The Math Gene: How Mathematical Thinking Evolved and Why Numbers Are Like Gossip*. Basic Books New York.

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2013. “Lies in Disguise—an Experimental Study on Cheating.” *Journal of the European Economic Association* 11 (3). Oxford University Press: 525–47.

Kahneman, Daniel, and Dale T Miller. 1986. “Norm Theory: Comparing Reality to Its Alternatives.” *Psychological Review* 93 (2). American Psychological Association: 136.

Kajackaite, Agne. 2018. “Lying About Luck Versus Lying About Performance.” *Journal of Economic Behavior & Organization* 153. Elsevier: 194–99.

Mazar, Nina, On Amir, and Dan Ariely. 2008. “The Dishonesty of Honest People: A Theory of Self-Concept Maintenance.” *Journal of Marketing Research* 45 (6). American Marketing Association: 633–44.

Mazar, Nina, and Dan Ariely. 2006. “Dishonesty in Everyday Life and Its Policy Implications.” *Journal of Public Policy & Marketing* 25 (1). American Marketing Association: 117–26.

Rankin, Joan L, Roger H Bruning, and Vicky L Timme. 1994. “The Development of Beliefs About Spelling and Their Relationship to Spelling Performance.” *Applied Cognitive Psychology* 8 (3). Wiley Online Library: 213–32.

Shalvi, Shaul, Ori Eldar, and Yoella Bereby-Meyer. 2012. “Honesty Requires Time (and Lack of Justifications).” *Psychological Science* 23 (10). Sage Publications Sage CA: Los Angeles, CA: 1264–70.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal. 2015. “Self-Serving Justifications: Doing Wrong and Feeling Moral.” *Current Directions in Psychological Science* 24 (2). Sage Publications Sage CA: Los Angeles, CA: 125–30.

Uttal, David H. 1997. “Beliefs About Genetic Influences on Mathematics Achievement: A Cross-Cultural Comparison.” *Genetica* 99 (2-3). Springer: 165–72.