IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11798

A Method to Estimate Mean Lying Rates and Their Full Distribution

Ellen Garbarino
Robert Slonim
Marie Claire Villeval

SEPTEMBER 2018

DISCUSSION PAPER SERIES

IZA DP No. 11798

A Method to Estimate Mean Lying Rates and Their Full Distribution

Ellen Garbarino
University of Sydney Business School

Robert Slonim
University of Sydney and IZA

Marie Claire Villeval
Université Lyon, CNRS, GATE and IZA

SEPTEMBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

A Method to Estimate Mean Lying Rates and Their Full Distribution*

Studying the likelihood that individuals cheat requires a valid statistical measure of dishonesty. We develop an easy empirical method to measure and compare lying behavior within and across studies to correct for sampling errors. This method estimates the full distribution of lying when agents privately observe the outcome of a random process (e.g., die roll) and can misreport what they observed. It provides a precise estimate of the mean and confidence interval (offering lower and upper bounds on the proportion of people lying) over the full distribution, allowing for a vast range of statistical inferences not generally available with existing methods.

JEL Classification: C91, C81, D03

Keywords: dishonesty, lying, econometric estimation, sampling errors, experimental economics

Corresponding author:
Marie Claire Villeval
GATE (UMR5824)
93 Chemin des Mouilles
F-69130, Ecully
France
E-mail: villeval@gate.cnrs.fr

* We are grateful to Quentin Thevenet for assistance. Financial support from the University of Sydney and the FELIS program of the French National Agency for Research (ANR14-CE28-0010-01) is gratefully acknowledged. This research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).

1. Introduction

In the last decade, dishonesty has received extensive attention in the economics literature, with major contributions from experimental studies. In experiments examining dishonesty, researchers commonly use techniques involving self-reports of privately observed outcomes from random devices, such as dice rolls (*e.g.*, Fischbacher and Föllmi-Heusi 2013; Shalvi *et al*. 2011) or coin tosses (*e.g*., Cohn et al., 2014). A recent systematic review by Abeler *et al*. (2016) identified more than 90 experimental papers dealing with dishonesty using subject's self-reports of privately observed outcomes of a random device.

In these studies, since the experimenter does not know the true outcome that the subject observes, it is not possible to measure whether any individual subject's report is honest or dishonest. Instead, researchers focus on population-level dishonesty by comparing the proportion of each reported outcome from the subject population with the expected proportion of each outcome that would occur if all subjects truthfully reported what they observed, based on the random device's actual distribution. This approach indicates whether the proportion of individuals reporting any particular outcome statistically differs from the expected proportion if all individuals reported honestly. However, it relies on the assumption that the fraction of subjects who observe the high payoff outcome is equal to the expected frequency of this outcome if subjects reported truthfully, which is thus subject to sampling errors and can be problematic especially in the case of small samples. Moreover, this approach does not provide the distribution of the proportion of subjects that behaved dishonestly. It is thus difficult to compare the extent of lying across treatments or studies beyond indicators for when dishonesty does and does not significantly occur.

To address these limitations, this paper provides an econometric technique to correct for sampling errors and to estimate the full distribution on the proportion of individuals who make dishonest reports (*i.e.*, lie). Our method estimates the proportion of liars using a weighted average of the estimated fraction of liars for any possible realization of the fraction of subjects who observed the high outcome, instead of using simply the expected frequency of the high outcome. To estimate the full distribution, the technique uses the information on the distribution of all possible outcomes the population can observe to infer the PDF and CDF of dishonesty. By estimating the distribution, our technique provides a precise estimate not only of the mean and confidence intervals (indicating lower and upper bounds on the proportion of people lying), but also allows for any other statistical inference that can be inferred from the full distribution. Monte-Carlo simulations used to compare our method to the existing methods indicate that our method makes fewer Type 1 errors (incorrectly not rejecting that no subjects are lying) than the other methods when the proportion of liars is small or medium. These simulations reveal incidentally that none of these methods provide a good approach for avoiding Type 1 errors with small samples, and thus suggests that large samples are necessary to insure sufficient power to reject Type 1 errors.

We have developed a software to implement our technique that is freely available at this address: http://lyingcalculator.gate.cnrs.fr. This software not only allows researchers to estimate the distribution of lying for their data, but also provides a trivial method to perform power calculations to determine appropriate sample sizes for new treatments. The technique presented in this paper, and the software, is easy to use and requires minimal information: the total number of subjects, the number of subjects reporting the high payoff and the probability (based on the random device) of receiving the low payoff.

The paper is organized as follows. Section 2 explains why we need a more precise technique than those typically used in the literature to estimate mean lying rates. Section 3 develops our econometric technique and Section 4 illustrates it. Section 5 compares estimates using our method and those based on other techniques. Section 6 concludes.

2. Techniques typically used in the literature: A simple demonstration of sampling errors

A typical method for estimating lying rates, for example in a coin toss task, assumes that the population consists of two types: honest individuals who report the winning outcome (for example, h=heads) with probability 0.5 and dishonest individuals who report this outcome with probability 1. The percentage of winning outcomes h reported by the subjects is written as follows: $h= l*1 + (1-l)*0.5 = 0.5*(1+l)$, where l *is* the proportion of dishonest subjects. Thus, given the proportion of subjects that report the winning outcome h, this formula implies that the proportion of subjects who lie is thus: $l = 2*h -1$. For example, if four subjects toss a coin and three report heads (i.e., h=3/4), the formula indicates that 50% lied; if two subjects report heads (i.e., h=1/2), it indicates that no one lied. This formula is used in *e.g.*, Houser *et al*. (2012) and in Cohn *et al*. (2015). Abeler *et al*. (2016) advance the analysis by using a general formula to report a proxy for the mean lying percentage. Their formula indicates that the percentage of subjects that lie equals $(r-q)/(1-q)$ where r is the percentage of subjects that report the higher payoff and q is the likelihood of observing this good outcome. For instance, with a coin toss (q=50%), if r=60%, then this formula implies that on average 20% of the subjects {(0.6-0.5)/0.5} would have lied. Note that these different formulas (as well as that proposed in Moshagen and Hilbig, 2017) are identical when the binary variable is distributed uniformly.

There are, however, limitations to these formulas. First, the estimate of the mean percentage may be biased because of the assumption that the fraction of subjects who actually observe the high outcome equals the expected frequency of the outcome of the random variable. This can most easily be seen if we look at when h (in the first formula) or r (in the second formula) equals

50%, in which case these formulas indicate that on average no subjects had lied. However, this ignores the distribution of possible outcomes that the subjects could have observed, which includes less than 50% observing the higher payoff outcome but reporting dishonestly. In the previous example with four subjects, the expected frequency of the high outcome is 50% but in fact there is 37.5% chance that two out of four subjects observed of the high outcome, 25% chance that one or three subjects observed it, and 6.25% chance that either no or four subjects observed it. Thus, there is a 25% chance that just one subject observed the high outcome, and hence 1 of the three subjects observed the low outcome and lied (i.e., a 33% lying rate) and a 6.25% chance that no subject observed the high outcome and hence 2 of the four subjects observed the low outcome and lied (i.e., a 50% lying rate). From that respect, sampling errors are problematic and especially in the case of small samples. Second, these formulas do not provide the distribution or any other statistics, such as confidence intervals, on the proportion that lied.

Our approach proposes a more precise estimate by correcting for sampling errors and by providing the entire CDF and PDF of lying including the correct mean of the distribution across the full range of possible outcomes. It uses a weighted average of the estimated fraction of liars for any possible realization of the fraction of subjects who observed the high outcome. In the previous example, it computes the difference between the frequency of the reported high outcome and the frequency that would be expected in the case of no, one, two, three or four subjects who actually observed the high outcome, weighted by the probability that each realization occurred. The method is explained more formally in the next section.

3. Estimating the full distribution of dishonesty in experiments

This section presents our technique to estimate the full distribution of the percentage of subjects lying. We first consider the case of two payoffs and then show how one can derive a binary statistic to apply our method for the case of $m > 2$ payoffs.

3.1 Two payoffs

Consider N subjects who privately observe the outcome of a random device (*e.g.*, die rolls, coin tosses) and report one of two possible payoffs: x_1 and x_2 ($x_1 < x_2$) that map directly from the random device (*e.g.*, Tails pays \$0, Heads pays \$5; 1-5 on the die pays \$0, 6 pays \$2). Outcomes x_1 and x_2 occur with probabilities p and 1-p, respectively. If R subjects ($0 \leq R \leq N$) report the high outcome x_2, we want to know the probability distribution of the percentage of subjects who observed the low outcome x_1 but reported dishonestly the high outcome x_2.

As in all previous methods used in the literature, we assume that players do not lie downward: any subject who reports the low outcome x_1 reports truthfully.[1] Letting T be the (unknown to the researcher) number of subjects that observed the high outcome, there are R+1 different possible numbers of subjects that could have observed the high outcome (*i.e.*, $0 \leq T \leq R$). For each possible T, the percentage of subjects who lied is:

$$L_{T,R,N} = [R–T] / [N–T] \qquad (1)$$

where the numerator is the difference between the number of subjects that reported the high outcome and the number of subjects that observed the high outcome (*i.e.*, the number who were dishonest), and the denominator is the difference between the number of subjects in the experiment and the number that observed the high outcome (*i.e.*, the number who could have lied). This equation is similar to that used in Abeler *et al*. (2016).

[1] Rigdon and d'Esterre (2015) were able to observe the true outcome subjects observed. In personal correspondence, the authors informed us that not a single subject lied when reporting the lower monetary outcome. Using a Bluetooth enabled die that covertly transmits the observed outcome of each roll to the experimenter Kröll and Rustagi (2016) were also able to observe the prevalence of underreporting. In personal communication, they reported that underreporting occurred in only 9 cases out of 2880 (0.3%). It is possible to relax our assumption of no underreporting (*e.g.*, we could allow for decision error so that with probability εj j reports an outcome that a subject did not intend to report). We leave this for future research not only since we expect decision errors are rare given the simplicity of the task, but also since we expect decision errors would be symmetric and thus only minimally affect the implications on our analysis.

To examine the probability of each possible realization of T, note that $Q_{T,N,p}$ = [(N! / (T!*(N-T)!)] * p^T * $(1\text{-}p)^{(N\text{-}T)}$ is the *unconditional* probability that T subjects would observe the high outcome if we allowed all possible observations for T from T = 0 to N. However, since we assume no subject is dishonest to report the low outcome if he observed the high one, there are at least N-R subjects that observed the low outcome, and thus the unobserved number of subjects T that observed the high outcome cannot exceed R. Thus, the probability that T subjects would have observed the high outcome when R subjects reported the high outcome is:

$$Z_{T,N,R,p} = \frac{Q_{T,N,p}}{\sum_{k=0}^{R} Q_{k,N,p}} \quad \text{for all T = 0 to R} \qquad (2)$$

where the numerator is the probability of the realization of T based on the binomial distribution with N observations, R successes and probability *p*. The denominator adjusts up these probabilities by the cumulative likelihood that between 0 and R subjects actually observed the high outcome.[2]

The PDF that $y = L_{T,R,N}$ subjects were dishonest when R subjects reported the high outcome is:

$$\Pr(y = L_{T,R,N}) = Z_{T,N,R,p} \qquad (3)$$

The CDF that $L_{T,R,N}$ or more subjects lied follows directly by summing over the PDF:

$$\Pr(y \geq L_{T,R,N}) = \sum_{k=0}^{T} (Z_{k,R,N,p}) \qquad (4)$$

The expected percent of subjects who lied, $E_{N,R,p}$, is the sum (over all possible observations T given R subjects reported the high outcome) of the percent of subjects who were dishonest weighted by the probability that T subjects observed the low outcome:

$$E_{N,R,p} = \sum_{T=0}^{R} (Z_{T,R,N,p} * L_{T,R,N}) \qquad (5)$$

[2] In personal correspondence, David Hugh-Jones notes that our adjustment of the denominator can over-estimate the percent of lying when the sample is small and when the true percent of subjects who lie is also small. A Bayesian approach and Maximum Likelihood estimation may correct this bias, but is beyond the scope of this paper.

Finally, the lower bound for the Confidence Interval (CI, hereafter) (p-value) that y subjects were dishonest is simply the minimum value of $L_{T,R,N}$ such that:

$$p/2 < \Pr(y \geq L_{T,R,N}), \tag{6}$$

and the upper bound for the CI (p-value) that y subjects were dishonest is simply the maximum value of $L_{T,R,N}$ such that:

$$\Pr(y \geq L_{T,R,N}) < 1\text{-}p/2 \tag{7}$$

3.2 More than two payoffs

In some experiments there are more than two payoffs (*e.g.*, across die outcomes or multiple coin tosses). Consider m distinct outcomes $x_1, \ldots, x_m$ with probability $p_1, \ldots, p_m$, respectively, with $x_1 < x_2 < \ldots < x_m$. The question we address is what percent of subjects who observe an outcome below some threshold x_k ($1 < k \leq m$) are dishonest and report an outcome equal to or above k.[3] To examine the expected percent of subjects who act dishonestly to avoid the worst k-1 outcomes to obtain one of the best m-k outcomes, let $p=p_1 + \ldots + p_{k-1}$ and then follow the procedure for two outcomes.

Bifurcating the data into two payoff groups might be interesting in a variety of cases. For instance, if payoffs are similar for the lowest k-1 outcomes and distinctly higher and similar for the top m-k outcomes, or to study whether subjects want to either avoid the lowest possible outcome (k=2) or obtain the highest possible outcome (k=m).

3.3 Implementation

Using the full distribution of possible outcomes, we can directly calculate the expected percentage of individuals who lied conditional on having observed the low payoff outcome(s)

[3] Researchers may also be interested in partial lying (*e.g.*, to examine how many subjects lie from one outcome x_k to x_j ($k < j$)) to measure small *vs.* big lies. We believe the techniques to estimate partial lies can be developed from the method presented in this paper, but leave it for future work as it requires technical sophistication beyond the scope of this paper.

($E_{N,R,p}$), and the PDF ($Z_{T,N,R,p}$) and CDF ($\Pr(y \geq L_{T,R,N})$). Using the CDF, we can also estimate the confidence interval for any level of significance to determine the minimum and maximum percentages y who lied. Thus, we are able to examine not only whether subjects were dishonest, but also (a) the mean expected percent who lied, (b) the minimum and maximum percentages we can be confident lied (*i.e.*, the lower and upper bounds on the CI), and (c) using the CDFs, whether two (or more) treatments significantly differ for any statistic of interest including for instance at the means, medians, quantiles and over the full distribution, using non-parametric tests.

4. Illustrations of the method of estimation of the percentage of lying

We now provide a few examples using our technique and online calculator. We only need to enter the total number of subjects, N, the number of subjects reporting the high outcome, R, the probability of observing the low outcome, p, and the desired CI. Figure 1 illustrates the interface of our software.

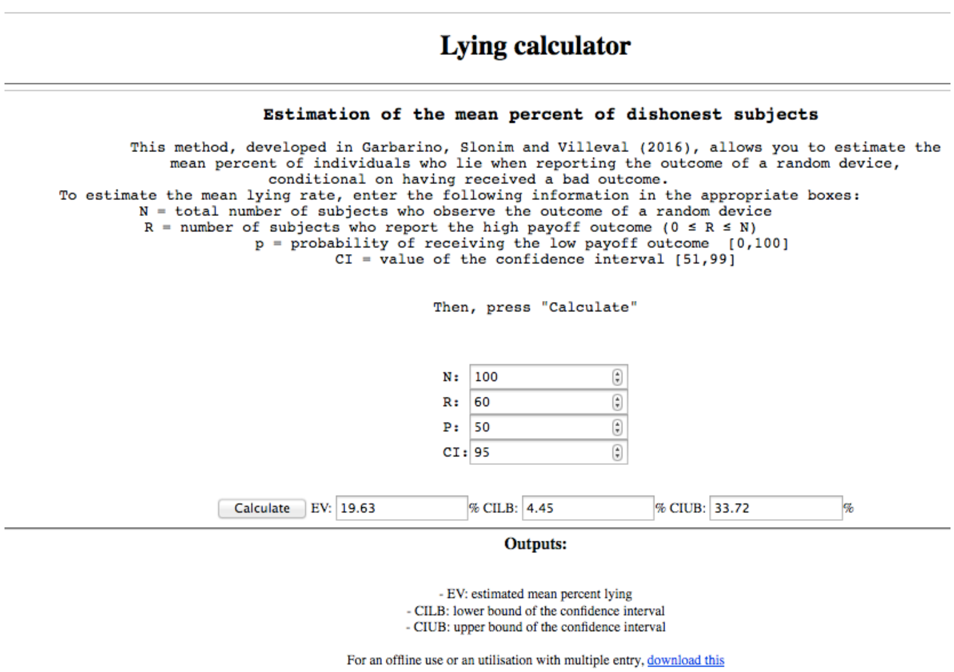


Figure 1. Calculator (http://lyingcalculator.gate.cnrs.fr)

Table 1 reports examples using N=100 and R=60 and shows results for various values of *p* and of the three most common CIs (*i.e.*, 90%, 95% and 99%). When *p* = 50%, and R=60 out of the N=100 subjects report the high payoff outcome, we can reject that all subjects were honest at the 90th and 95th percent CI, but not at the 99th percent. At the 90th, 95th and 99th percent CI, we can be confident that at least 7.0%, 4.4% and 0.0% of subjects lied, respectively, and at most 32.0%, 33.7% and 36.9% lied. If the likelihood of observing the low outcome was higher, rising from *p* = 50% to 60%, to 90%, to 97%, we would be 95% confident that at least 4.4%, 21.2%, 52.5%, and 57.3% lied, respectively, and at most 33.7%, 42.8%, 58.3%, and 60% lied.[4]

Table 1. Illustration of the method of estimation of the percentage of lies

Example	Inputs				Outputs		
	N	R	*p*	CI	EV	Lower bound	Upper bound
1	100	60	50%	95%	19.63%	4.44%	33.72%
2	100	60	50%	90%	19.63%	6.98%	31.95%
3	100	60	50%	99%	19.63%	0.00%	36.91%
4	100	60	40%	95%	7.88%	0.00%	22.31%
5	100	60	60%	95%	32.88%	21.25%	42.85%
6	100	60	90%	95%	55.51%	52.51%	58.32%
7	100	60	97%	95%	60.00%	57.31%	60.00%

Notes: N = number of subjects; R = number who report the higher outcome; *p* = probability of the lower outcome; CI = confidence interval; EV = expected value; lower and upper bound of the CI.

The calculator can also be easily used to determine the power of a study. Suppose a researcher wishes to determine the power to detect lying in a one shot coin toss study if he recruits N=100 subjects and expects that the likelihood each subject will lie is 20% whenever the subject observes the bad outcome. In Table 2, we show the 95% confidence interval for a few of the possible outcomes R (the number of subjects that report the good outcome). To obtain 95%

[4] Given R=60 subjects out of the N=100 subjects reported the higher outcome, we assume that the highest possible percent who could have been dishonest is 60%.

confidence, the power of the experiment will depend on how likely we will observe R=58 or more subjects who will report the good outcome.

Table 2. Illustration of method to determine power

Example	Inputs				Outputs		
	N	R	*p*	CI	EV	Lower bound	Upper bound
1	100	53	50%	95%	9.29%	0.00%	22.79%
2	100	54	50%	95%	10.43%	0.00%	24.27%
3	100	55	50%	95%	11.68%	0.00%	25.77%
4	100	56	50%	95%	13.05%	0.00%	27.32%
5	100	57	50%	95%	14.55%	0.00%	28.89%
6	100	58	50%	95%	16.15%	2.44%	30.48%
7	100	59	50%	95%	17.85%	3.24%	32.10%

Notes: N = number of subjects; R = number who report the higher outcome; *p* = probability of the lower outcome; CI = confidence interval; EV = expected value; lower and upper bound of the CI.

To calculate the power of this study, we simulate the experiment where each of N=100 simulated subjects has a 50% (independent and randomly drawn) chance to observe the good or bad outcome (*i.e.*, Heads or Tails). If the simulated subject observes the good outcome, he honestly reports it. But if he observes the bad outcome, there is a 20% random chance (based on the researcher's expectation) that he will report the good outcome, and an 80% chance he will honestly report the bad outcome. Then, we sum up the total number of subjects that report the good outcome to get R^s. We run this simulated experiment S times (s=1 to S) to determine the distribution of R assuming N=100, *p*=50%, and the experimenter's belief that a subject is 20% likely to lie if he observes the bad outcome.

We ran this simulation experiment S=10,000 times. The results indicate that in 90% of the simulations R ≥ 54, in 85% of them R ≥ 55, in 80% of them R ≥ 56, in 75% of them R ≥ 57 and in 69% of them R ≥ 58. Given that by inspection of Table 2 we need to have R=58 or more subjects report the good outcome, and based on the simulations with the researcher's expectation, the researcher will have 69% power to detect lying at the 95% confidence level.

5. A comparison with other methods

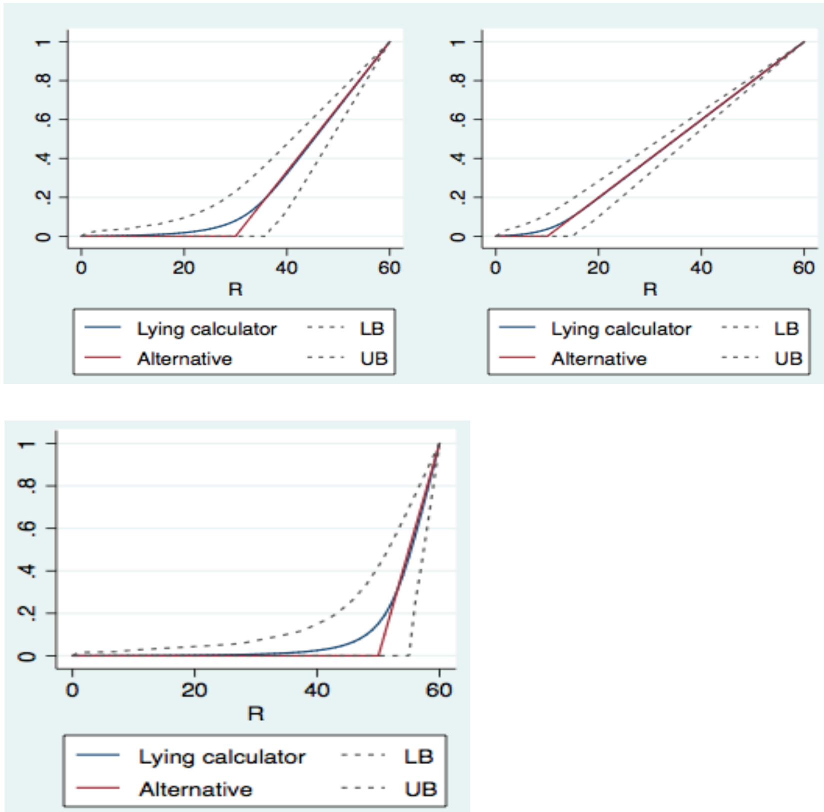
In this section we show evidence of how the estimates of our approach differ from estimates using other methods, discussing the advantages and limitations of this approach. Its main advantages over other methods are that it provides a more precise estimate of the proportion of liars in small samples and a confidence interval, and it identifies a fraction of liars where other studies would not have been able to detect them.[5]

More systematically, figure 2 compares the measures using our lying calculator with the method in Abeler *et al*. (2016) for three examples.[6] Recall that in their method, the estimated percentage of subjects that lie equals $(r\text{-}q)/(1\text{-}q)$ where r is the percentage of subjects that report the higher outcome and q is the likelihood of observing this good outcome. We have run the analysis with N=60 and with N=600 subjects. The x-axis represents the number of subjects R who report the high outcome and the y-axis represents the estimate of the mean percentage of lying. The top left panel corresponds to a coin toss with 50% chance to observe the low outcome. The top right panel corresponds to a die roll with 5/6 chance to observe the low outcome. The bottom panel corresponds to a die roll with 1/6 chance to observe the low outcome. Both methods provide estimates of the mean, but our method has the advantage of providing the full distribution, and hence the confidence intervals shown in Figure 2.

N=60

[5] Keeping the same example as presented in Section 2, consider the case where only one out of four subjects report the high outcome after tossing a coin. Other methods conclude that there is no lying (assuming no downward lying). Our method accounts for the fact that it may have occurred that no subject actually observed the high outcome (with a 6.25% probability) or that only one subject observed the high outcome, and thus nobody lied. Our method predicts a proportion of liars of 5% (a 6.25% chance that 25% of the subjects lied plus a 25% chance that no subjects lied) with a 95% confidence interval from 0 to 25%.

[6] Note also that Abeler *et al*.'s (2016) approach would indicate 20% lying in examples 1-3 of Table 1, and 0%, 33.3%, 55.6% and 58.8% for examples 4 through 7. Thus, their formula gives a close estimate in six of the seven cases, but is off by 8% in example 4.



N=600

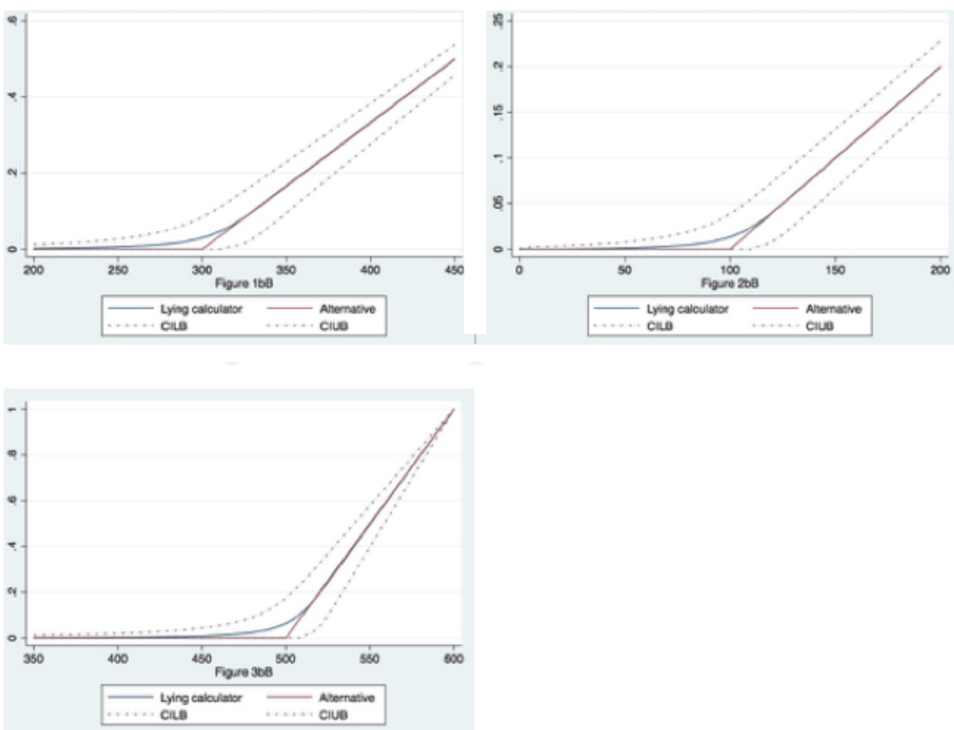


Figure 2. Comparisons between our method and the method proposed in Abeler *et al.* (2016) for N=60 and N=600

Note: CILB for lower bound of the confidence interval, CIUB for upper bound of this interval. The continuous lines represent the estimated lying rate using our calculator or based on Abeler et *al.*'s (2016) approach. The dashed lines represent confidence intervals. The top left panel corresponds to a coin toss with 50% chance to observe the low outcome, the top right panel to a die roll with 5/6 chance to observe the low outcome, and the bottom panel to a die roll with 1/6 chance to observe the low outcome.

The main differences are concentrated around the range of observations close to the expected outcome of the random device (for example, when R – the number of subjects reporting the good outcome - is between 20 and 35 for the case in which N=60, and the probability of the low outcome is 0.5; see the top left panel). Figure 2 indicates that the problem is more serious when the sample size is smaller. Other methods may fairly accurately estimate the actual lying rate when there is 10 percent or more lying but they may over-estimate differences in lying between treatments when one treatment has close to zero lying and another treatment has higher lying rates. For instance, using a Fisher exact test, Jacobsen and Piovesan (2016) report a marginally significant difference in lying when reporting a "6" in a die task between their Tax treatment and their Baseline, whereas our method finds a 90 percent confidence interval for lying of 0.0% to 7.3% in the Baseline and 0.0% to 17.2% in the Tax treatment, which indicates no significant difference between the treatments. Similarly, we applied our method to the data from Cohn *et al.* (2014) in which subjects have to toss a coin ten times, defining a bad outcome as having at most five winning coins. The estimated mean lying rate (reporting six or more winning coins when observing at most five) is 5.9% in the control condition and 25.7% in the condition where identity was primed. However, the significance test fails if one considers the overlap between the 95% confidence intervals (0-17.4 in the control and 10.1-38.9 in the priming condition).[7]

To further highlight the relative effectiveness of our method compared to the existing methods, we also report results of Monte Carlo simulations examining Type 1 and Type 2 errors.

[7] It remains that, as shown by Cohn *et al.* (2014), the success rate reported by the bankers whose identity was primed was significantly higher than the success rate reported by the control group.

We again consider a simple coin flip experiment where reporting Heads provides a higher payoff than reporting Tails, subjects privately observe the outcome of the coin toss and we assume, as other methods do, no downward lying so that subjects honestly report Heads if they observe Heads. For each simulation, let N be the number of subjects and $Q > 0\%$ be the (unobserved) true probability that each subject will lie if she observes Tails (*i.e.*, if a subject observes Tails, then she reports Heads with probability Q and Tails with probability 1-Q). A single simulation consists of (a) one coin flip for each subject with equal probability of observing Heads or Tails, (b) if the subject observes Heads, then she reports Heads, and (c) if the subject observes Tails, then we randomly determine if she reports Heads (*i.e.*, lies) with probability Q or she reports Tails with probability 1-Q. For each simulation, we count up the number of subjects who report Heads, R, and then determine three statistics:

(1) *Type 1 error for existing methods:* using the existing approaches that compare R to the binomial distribution (*e.g.*, Abeler *et al*. 2016), we determine whether the simulated data will not be able to reject the null hypothesis of 'no lying' (that is, whether the observed number of Heads is small enough that we cannot reject 0% lying; whenever we cannot reject the null hypothesis of 'no lying', we commit a Type 1 error by incorrectly not rejecting that no subjects are lying.

(2) *Type 1 error for our proposed method:* based on the number of reported Heads, R, we determine whether 0% lying is within the 95th percent confidence interval that our method generates; if 0% is within the confidence interval, then we commit a Type 1 error by incorrectly not rejecting that no subjects are lying.

(3) *Type 2 error for our proposed method:* based on the number of reported Heads, R, we determine whether Q% lying is outside (below or above) the 95th percent confidence interval that our method generates; if Q% is outside of the confidence interval, then we have made a Type 2 error.[8]

We ran S = 10,000 simulated experiments each for every combination of values of N = 20, 40, 60 and 100 (*i.e.*, experiments with 20, 40, 60 and 100 subjects) and values of the true lying rates of Q = 10%, 25%, 50% and 75% (*i.e.*, the likelihood that each subject lies if she observes

[8] We are not able to include Type 2 error calculations for the existing methods since, as noted previously, the existing methods do not provide any distribution to determine confidence intervals beyond the expected mean lying rates, thus there is no basis to determine whether these methods' estimates of the lying rate are (or are not) close to the true lying rates that generated the data.

the lower payoff outcome of the coin flip). Table 3 reports the results of the simulations. The first two columns indicate the simulated experimental conditions for the true lying percent *Q* of subjects who lie if they observe the lower payoff outcome (Column 1) and the number of subjects in the experiment (Column 2). Columns 3-5 present the results. Columns 3 and 4 show the percent of Type 1 errors when using the existing methods and when using our method, respectively. Column 5 indicates the percent of Type 2 errors for our method.

Table 3. Monte Carlo Simulations of Relative Effectiveness

Experimental Conditions		**Results**			
		Existing methods	Proposed method		
Q Prob a subject lies if she observes the bad outcome	*N* Number of subjects in each experiment	Type 1 Error: Prob reject no lies	Type 1 Error: Prob reject no lies	Type 2 Error: Prob reject true lying *Q*	Range of Heads reported where we do not reject *Q*
10%	20	94.3%	94.3%	5.7%	3 to 14
	40	92.7%	87.0%	7.5%	11 to 26
	60	92.5%	81.8%	4.5%	21 to 39
	100	87.1%	70.2%	4.4%	41 to 63
25%	20	82.2%	82.2%	11.3%	9 to 15
	40	68.3%	55.5%	10.8%	20 to 29
	60	60.0%	38.7%	12.1%	31 to 42
	100	34.0%	10.7%	12.5%	55 to 64
50%	20	37.9%	37.9%	31.4%	13 to 16
	40	10.4%	5.5%	28.3%	27 to 32
	60	2.8%	0.7%	24.1%	41 to 48
	100	0.1%	0.0%	25.2%	70 to 79
75%	20	3.1%	3.1%	50.5%	17 to 18
	40	0.0%	0.0%	47.8%	34 to 36
	60	0.0%	0.0%	43.8%	51 to 54
	100	0.0%	0.0%	36.6%	85 to 89

The results are the following. As should be the case, both methods are less likely to make Type 1 errors (*i.e.*, to not reject ‘no lying’) as either the sample size increases or as the true lying percent of the subjects increases. For instance, when the likelihood that each subject lies is

Q=50%, then the existing methods will make a Type 1 error 37.9%, 10.4%, 2.8% and only 0.1% of the time as the sample number of subjects increases from 20 to 40 to 60 to 100, respectively. Column 3 also shows that when an experiment has N=60 subjects, then the existing methods will make a Type 1 error 92.5%, 60.0%, 2.8% and 0.0% of the time as the percent of subjects who will lie Q (if they get the bad outcome) increases from 10% to 25% to 50% to 75%.

The simulations show that when we compare our proposed method to the existing methods, our method makes either the same or fewer Type 1 errors. We make the same amount of Type 1 errors when either the sample size is small (N=20) or when lying is extremely prevalent (Q=75%). However, for experiments with 40 to 100 subjects and 10%, 25% or 50% of subjects who will lie if they get the bad outcome, our method produces consistently fewer Type 1 errors. For instance, an experiment with N=100 subjects where Q=25% of subjects will lie if they get the bad outcome, the existing methods will make a Type 1 error 34.0% of the time whereas our proposed method will make a Type 1 error only 10.7% of the time.

The simulations results for Type 1 errors also provide an important cautionary message for researchers investigating lying behavior. For the lying rates we explored in the simulations, no method provides a good approach for avoiding Type 1 errors with smaller samples, and thus suggests large samples will be necessary to insure sufficient power to reject 'no lying'. The simulation results show that experiments that have found 'no lying' with samples even up to 100 subjects, may be highly prone to have made a Type 1 error. For instance, using the existing methods, the simulations indicate that studies with 100 or fewer subjects that have found 'no lying' are at least 34% likely to have made a Type 1 error even when 25% of the subjects would have lied if they had observed the bad outcome and almost 90% likely to have made a Type 1 error even with 100 subjects if just 10% of the subjects would lie. While our improved method helps reduce Type 1 errors, the results show that we still would need substantially more than 100 subjects to have the power at conventional levels, such as 80% power (*i.e.*, 20% or less chance of making a Type 1 error), to detect lower levels of lying such as 10%.

The fifth Column of Table 3 shows the probability of our method making a Type 2 error in which the true lying probability in the subject population is outside the 95th percent confidence interval. It indicates, as expected, that the likelihood of Type 2 errors generally decreases as the sample size increases, *ceteris paribus*. For instance, the percent of Type 2 errors that occur when Q=50% of the subjects would lie if they observed the low payoff outcome falls from 31.4% to 28.3% to 24.1% as the number of subjects increases from 20 to 40 to 60.[9] The results in Column 5 also show that Type 2 errors increase as the true percent of subjects who lie, Q, increases, all else equal. For instance, for N=100 subjects, the percent of Type 2 errors increases from 4.4% to 12.5% to 25.2% to 36.6% for Q increasing from 10% to 25% to 50% to 75% lying, respectively. This pattern occurs because the range of the high payoff reports, R, that would allow us to avoid a Type 2 error is shrinking the higher the reported number of the higher payoff outcomes. For instance, when Q=10% and N=100, any total number of reports of the good outcome, R, from an experiment in the range from 41 to 63 would let us avoid making a Type 2 error (see Column 6), whereas only reported values of good outcomes of R from 85 to 89 would let us avoid making a Type 2 error when Q=75% and N=100. Future work might explore whether the incorporation of Bayesian prior beliefs (when available) could further reduce the incidence of Type 2 errors.

Overall, the comparison of the results in Columns 3 and 4 shows that our method reduces Type 1 errors compared to existing methods, and with reasonable sample sizes, we can mostly avoid Type 1 errors. Column 5 demonstrates that with our method (there is no equivalent with

[9] In further simulations not shown here, if the sample size increases to N=1,000 subjects, the percent of Type 2 errors further decreases for all levels of Q explored here, but this does not appear to be of practical interest since most research budgets would not extend to that many subjects.

existing methods) Type 2 errors are low when the true probability of subjects lying is low, but when there are substantially high rates of lying, we should be more cautious in assuming we are estimating the true lying rate (though we can confidently reject 'no lying').

6. Discussion

With only minimal information, our lying calculator provides a precise estimate of the mean percentage of lying and confidence intervals indicating lower and upper bounds on the proportion of people lying. These estimates correct for sample errors and allow for a range of statistical inferences to be made allowing for the comparison of levels of lying across treatments or studies, which is useful for meta-analyses of the literature (*e.g.*, Garbarino *et al*. 2017), or for comparing treatments or sub-groups of the population (*e.g.*, Banerjee *et al*., 2017; Benndorf *et al.* 2017). The technique uses the information on the full distribution of possible outcomes that the individual could have observed to infer the PDF and CDF of lying. Given the ease of use for the lying calculator and its ability to give more powerful and more accurate estimates for statistical analyses, we strongly recommend the use of this technique in future research exploring the role of dishonesty in decision making.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2016). Preferences for truth-telling. CEDEX Discussion Paper No 2016-13, University of Nottingham.

Banerjee, Ritwik, Nabanita Data Gupta and Marie Claire Villeval (2017). The spillover effects of Affirmative Action on Competitiveness and Unethical Behavior. *European Economic Review* 101: 567-604.

Benndorf, Volker, Claudia Moellers, and Hans-Theo Normann (2017). Experienced vs. inexperienced participants in the lab: Do they behave differently? *Journal of the Economic Science Association* 3.1: 12-25.

Cohn, Alain, Michel André Maréchal, and Thomas Noll (2015). Bad Boys: How Criminal Identity Salience Affects Rule Violation. *Review of Economic Studies*, 82 (4): 1289–1308.

Fischbacher, Urs, and Franziska Föllmi-Heusi (2013). Lies in disguise: An experimental study on cheating. *Journal of the European Economic Association* 11(3): 525–547.

Garbarino, Ellen, Robert Slonim, and Marie Claire Villeval (2017). Loss aversion and lying behavior: Theory, Estimation and Empirical Evidence (July 2017). GATE Working Paper no 1631. Available at SSRN: https://ssrn.com/abstract=2876636.

Houser, Daniel, Stefan Vetter and Joachim Winter (2012). Fairness and Cheating. *European Economic Review* 56: 1645-1655.

Jacobsen, Catrine, and Marco Piovesan (2016). Tax me if you can: An artefactual field experiment on dishonesty. *Journal of Economic Behavior & Organization* 124: 7-14.

Kröll, Markus, and Devesh Rustagi (2016). Shades of Dishonesty and Cheating in Informal Milk Markets in India. SAFE Working Paper, 134, Goethe University, Frankfurt.

Moshagen, Morten, and Benjamin E. Hilbig (2017). The statistical analysis of cheating paradigms. *Behavioral Research* 49: 724-732.

Rigdon, Mary L., and Alexander D'Esterre (2015). The effect of competition on the nature of cheating behavior, *Southern Economic Journal* 81 (4): 1012-1024.

Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten K.W. De Dreu (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* 115: 181–190.