Measuring Individual Dishonesty Using (Repeated) Probabilistic Lying Games

Nikolaj A. Harmon
University of Copenhagen

June 2021

Abstract

Much recent research studies the relationship between people's inherent propensity for dishonesty and various economic outcomes. To measure individual dishonesty, researchers rely on experimental games in which individuals can gain financially by lying about the outcome of one or more die rolls, coin tosses or other random outcomes. The way behavior in these games is translated into quantitative measures of dishonesty is ad hoc and varies from paper to paper. In this paper, I present a simple econometric framework formalizing probabilistic lying games and how behavor in these games relate to individual differences in dishonesty. I use this to discuss different formal definitions of dishonesty and to compare different existing dishonesty measures from the literature. I show how data from probabilistic lying games can be used to construct unbiased to different formal measures of individual dishonesty. These measures are subject to classical measurement error, however, leading to attenuation bias problems when employed in regression analyses. When each individual participates in more than one round of lying games however, a simple split sample instrument is available that addresses measurement error problems. I also discuss possible extensions to the case where researchers observe the actual outcome of the die rolls, coin tosses or other outcomes in the game. Finally, I provide a set of practical recommendations for applied researcehers.

Keywords: honesty, die-under-cup, coin-tossing, population inferred cheating task

1 Introduction

When an individual is deciding whether to lie, accept a bribe or undertake some other dishonest action, their decision is likely to be influenced by a range of external factors such as the potential rewards from dishonest, the likelihood of being found out and the potential costs of this happening. Additionally, however, their final decision is likely to also depend on their own inherent propensity for dishonesty, reflecting either a fundamental personality trait or possibly other traits and characteristcis that shape their willingness to be dishonest. Such differences in inherent dishonesty has garnered increasing attention recently within economics and related fields. Specifically, there has been a growing interest in examining how individual dishonesty affects or relates to various

economic outcomes and other behaviors. This includes for example corruption and the willingness to work in the public sector (Barfort et al., 2019; Hanna and Wang, 2017; Gans-Morse et al., 2021; Gans-Morse, Forthcoming), school misconduct (Cohn and Maréchal, 2018) or fraudulent provision of lower quality goods to consumers (Kroell and Rustagi, 2021).

This recent interest in dishonesty has in part been inspired by methodological advances in the measurement of dishonesty. Recent papers build on the experimental approach to measuring dishonesty introduced by Fischbacher and Föllmi-Heusi (2013). The approach involves asking individuals to participate in one or more rounds of a game in which they have to report the outcome of some random process, such as a die roll or coin flip. Importantly, the game is designed so that only the participant ever observes the true outcome, while the outcome *reported* by the participant determine their winnings. This creates an opportunity for the participant to dishonestly raise their own winnings by lying about the true outcome without any risk of this being exposed. At the same time, however, the extent to which an individual reports high winning outcomes contains information about their likelihood and extent of lying. Games following in this mold are sometimes referred to as a form *of Population Inferred Cheating Tasks*, or as *die-under-cup* or *coin-tossing* games for the specific case of reporting die rolls or coin tosses (Jacobsen et al., 2018). Since this paper shall be concerned with the general class of games not necessarily relying on coin-tosses or dice rolls, and since this paper chiefly deals with individual rather than population-level dishonesty, I shall in this paper refer to the general class of games as *probabilistic lying games.*

The usefulness of probabilistic lying games for studying inherent dishonesty is intuitively obvious: the group of individuals who appears to have gotten especially lucky should 'on average' contain more dishonest individuals - especially if this is based on observing many rounds of games. Relying on this intuition most existing papers construct various intuitive measures of 'how lucky' an individual appear to have gotten in the game and use this as a proxy for individual dishonest in their analysis. The employed measures include the share or total number winning reports (e.g. Gans-Morse et al. (2021) and Cohn and Maréchal (2018)), the share of reported die rolls that are above some cutoff (e.g. Gächter and Schulz (2016)), the total or average number of reported dice points (e.g. Hanna and Wang (2017) and Gächter and Schulz (2016)), indicators for whether reported points or total winnings are above the median or quartile (e.g. Hanna and Wang (2017) and Potters and Stoop (2016)), indicators based on the p-value for whether the reported number of dice points is significantly different from what would be expected under full honesty (e.g. Kroell and Rustagi (2021)) or standardized versions of the reported outcome (e.g. Abeler et al. (2019)). While all these measures are intuitively reasonable, they suffer from at least two major drawbacks: i) Because it is often unclear how each measure relates to various definitions of dishonesty, it is often unclear how to think about the magnitudes of different estimates. Moreover, it is partciularly unclear how to generalize and contrast findings across studies, ii) The statisticial and econometrics properties of each of these measures are not clear. It is unclear exactly what assumptions underlie the empirical approach invoked. Morever, while many papers acknowledge that the employed measures of dishonesty must be subject to important measurement error, it differs wildly whether and

how this is dealt with in the empirical analysis.

In this paper, I make an attempt at remedying these drawbacks. To do this, I first present a simple econometric framework that describes probabilistic lying games and links behavior in these games to a very general concept of individual dishonesty. I then use this framework to discuss different empirically-relevant definitions of the individual propensity for dishonesty and to examine the properties of different empirical measures that can be constructed using data from a probabilistic lying game. I focus specifically on the situation in which a researcher has individual data from a probabilistic lying game as well as data on another variable of interest plus possibly some control variables. The researcher is interested in using linear regression to estimate the relationship between some scalar measure of inherent dishonesty and some other variable of interest. The objective may be to estimate a causal effect of inherent dishonesty on the other variable, it may be to estimate a causal effect of the other variable on dishonesty or it may simply be to descriptively characterize the relationship between the variables. This case covers all of the cited papers above.

Results from the analysis naturally break down probabilistic lying games into two groups: Binary lying games in which the observed outcome is binary (e.g. a coin toss) and non-binary lying games in which the observed outcome can take many values (e.g. a die roll).

For behavior in binary lying games, I show that all individual differences in dishonesty collapse naturally to a scalar measure, the individual *cheat rate*, which simply measures the probability that the individual decides to cheat in any given round of the game. Under standard assumptions, I show that it is straightforward to construct an unbiased estimate of the individual cheat rate using data from a Probabilistic Lying Game. Importantly, this is true regardless of how many rounds each individual is asked to play in the game, including if each individual only plays one round.

I further show that the estimated cheat rate is subject to measurement error from two sources: i) random variation in the actual outcomes causes some individuals to appear more or less dishonest than they actually are, and ii) random variation in behavior over time imply that individuals may end up engaging in more or less dishonest behavior in the lying game than they do on average. Importantly, I show that this combined measurement error is classical in the sense that it is uncorrelated with the other variables in the analysis under standard assumptions. As a result, regression analyses that use dishonesty as the outcome variable produce consistent estimates under the usual conditions. Analyses that use dishonest as the independent variable however will produce (asymptotically) biased estimates and suffer from the usual attenuation bias problem.

Next, I propose a simple way to address measurement error bias in analyses that use dishonesty as the independent variable. The approach requires a suitable restriction on time dependence in dishonest behavior and that each individual participates in at least two rounds of the probabilistic lying game. If these requirements are satisfied, however, I show that a simple split-sample instrument approach can be used to provide consistent estimates. The approach simply involves computing two separate measures of the cheat rate that use data from different rounds of the lying game, and then using one of the measures as an instrument for the other in a standard 2SLS estimator. The split-sample instrument also offers a straightforward way to estimate the variance of individual

dishonesty.

Finally, I compare the estimated cheat rate measure of dishonesty with some of the other measures used in the literature. Many turn out to be trivially linked to the estimated cheat rate by by simply being a linear transformation of it. This makes it straightforward to relate regression estimates from previous work to the cheat rate and the formal definition of dishonesty presented here.

Turning to non-binary lying games, I show that there is no single natural way to collapse differences in dishonest to a scalar measure for these agmes. This in part reflects that dishonest behavior in the non-binary case involves both an extensive margin decision of whether to lie about the oucome or not, as well as an intensive margin decision about how much to lie about the true outcome.

I argue that an attractive scalar definition of dishonesty still exists that captures both the intensive and extensive margin: The *average misreport* measures on average how much an individual overreports by. Using this definition of dishonesty, I show that data from a probabilistic lying game can be used to construct an unbiased estimate of individual dishonesty. Again this turns out to be possible even when individuals only participate in a single round of the lying game.

Additional results for non-binary case follow the binary case closely: The estimated average misreport turns out to be subject to measurement error from the same sources as the estimated cheat rate in the binary case. Again the measurement error turns out to be classical however. I also show how the split-sample instrument approach can be used to address the bias from measurement error when dishonesty is used as the independent variable in a regression, as well as to estimate the variance of dishonesty. Comparing the average misreport to measures used in the existing literature also shows them to be linear transformations of each other, meaning that estimates from different studies can be related to the formal definiton of dishonesty presented here.

Finally, I briefly discuss the possibility of separating the extensive and intensive margin of dishonest. The overall message here is negative, as there is no general way to separate intensive and extensive margin dishonesty using data from a non-binary lying game. I clarify in a formal sense however that binary lying games isolate the extensive margin component of dishonesty whereas non-binary lying games measure a combination of the extensive and intensive margins. This opens the possibility of combining binary and non-binary lying games to separate intensive and extensive margin dishonesty. I also briefly discuss the possible advantages of modifuing the experimental design of the lying games so that the true outcome is in fact observed by the researcher (as in Kroell and Rustagi (2021)). In addition to lowering (but not removing) the measurement error in dishonesty, this modification offers another promising way to separating intensive and extensive margin dishonesty.

This paper contributes to the literature on measuring individual's inherent dishonesty and relating it to other outcomes and characteristics. The paper also has several broader methodological ties however. The econometric framework that I use is closely related to the theoretical framework used by Abeler et al. (2019) to draw empirical implications of different theoretical models of lying.

The points made about measurement error and attenuation bias here are also related to the points raised by Moshagen and Hilbig (2017), although they only consider the case of a non-repeated lying game and thus do not consider the possiblity of using the split-sample instrument to address measurement error. The use of the cheat rate as scalar measure of dishonest was introduced originally in Barfort et al. (2019), which also presents a parametric approach to separating out measurement error from the distribution of dishonesty. Under additional assumptions, Barfort et al. (2019) also discusses how the design of the probabilitic lying game affects measurment error in individual dishonesty, specifically noting that measurement error is decreasing with the number of rounds in the game. Finally, by essentially relying on repeated measurements of dishonesty to addresse measurment, the split-sample instrument approach is highly related to the ORIV estimator of Gillen et al. (2019). Whereas Gillen et al. (2019) focus on the general case of measurement error in experimental measures and assume that this error is classical, this paper provides primitive assumptions on probabilistic lying games such that appropriately chosen measures can be shown to only suffer from classical measurement error. As discussed in the paper, however, the split sample instrument approach for probabilistic lying games may be combined with the ORIV estimator for a potential efficiency gain.

The layout of the rest of the paper is as follows: Section 2 presents the general econometric framework describing both binary and non-binary lying games. It also presents and discusses the maintained assumptions invoked throuthougt the analysis. Section 3 then presents the analysis and results for binary lying games, while Section 4 covers the non-binary case. Section 5 concludes and provides a set of practical recomendations.

2 Econometric framework

A researcher has data on N participants (individuals) indexed by i. For each participant, the researcher observes some variable of interest, Y_i, and (possibly) a vector of controls X_i, which includes a constant. In addition, each participant has some *inherent propensity for dishonesty*, D_i, which is not directly observed in the data. Instead, as described in the next subsection, the researcher observes the participant's behavior in a probabilistic lying game. A main part of this paper will be concerned with how to define and measure D_i based on observed behavior in the game.

I will focus mainly on the case where the researcher's aim is to characterize the relationship between the propensity for dishonest D_i and the variable of interest Y_i using a linear regression framework that possibly conditions on the controls in X_i. The researcher is thus interested in either the parameter β_{LHS} or β_{RHS} from one of the following regression models that either includes dishonesty on the right or left hand side:

$$Y_i = \beta^{RHS} D_i + X_i' \pi^{RHS} + \varepsilon_i^{RHS} \quad , \quad E[\varepsilon_i^{RHS} | D_i, X_i] = 0 \tag{1}$$

$$D_i = \beta^{LHS} Y_i + X_i' \pi^{LHS} + \varepsilon_i^{LHS} \quad , \quad E[\varepsilon_i^{LHS} | Y_i, X_i] = 0 \tag{2}$$

Whether β_{LHS} or β_{RHS} is the more relevant parameter of interest depends on the specifics of the setting and study. If the researcher has in mind a causal interpretation of the relationship between D_i and Y_i, the direction of the causal relationship determines which of the regression models is relevant. If instead the aim of the analysis is descriptive, either model may be relevant for summarizing the empirical relationship between D_i and Y_i. As I expand on later, classical measurement error in the measure of D_i may in these cases make model 2 more attractive.

2.1 The probabilistic lying game

Each of the participants in the data have participated in K rounds of a probabilistic lying game. I let t index the rounds of the game. In what follows, I will be modelling behavior in the game as a (possibly) probabilistic process. Variation in dishonesty across individuals will be reflected in the fact that this process differs across i. As a matter of notation, conditional probabilities and conditional expectations that condition on the identity of an individual i will therefore be denoted P_i and E_i.

In each round of the game, each participant first observes some random outcome, $w_{it} \in \Omega$, which under full honesty would determine their winnings for the round. For example, w_{it} may be the value of a die roll, may be an indicator for a coin coming up tails or may be an indicator for correctly guessing a random outcome. Consistent with these examples, I will assume that w_{it} is discrete, although most of the results below extend trivially to the continuous case. Finally, throughout most of what follows, w_{it} will be assumed to be unobserved to the researcher (in Section 4.5, I return briefly to the case when w_{it} is observed).

After seeing the outcome w_{it}, the participant makes a report of the outcome to the researcher which will determine their payoff for the round. I let, $r_{it} \in \Omega$ denote the reported outcome for round t from participant i. Without loss of generality, I will assume that the participant's payoff is increasing in r_{it}. Because participants are free to report any outcome regardless of what was actually observed, this creates a scope for participants to dishonestly increase their payoff through misreporting. I let $m_{it} \equiv r_{it} - w_{it}$ denote the misreporting of participant i in round t.

2.2 A general definition of the individual propensity for dishonesty

The premise of the empirical analysis is that some individuals are more likely to misreport in a given situation. Accordingly, I will assume that the likelihood of misreporting by various amount is different across individuals. For a given person, I will allow for the possibility of dependence in lying behavior over time, however, I will impose that the sequence of lying behavior satisfies a stationarity assumption. This reflects that in order for the concept of inherent propensity for dishonest to be well-defined, dishonest behavior need to exhibit some stability over time. I will let $F_i(\cdot|\cdot)$ denote

the cumulative distribution function for person i's lying behavior conditional on the outcome they observe in any given round of the lying game.

Assumption 1. $P_i(m_{it} \leq x | w_{it} = w) = F_i(x|w)$ *for all* t

The distribution for person i, $F_i(\cdot|\cdot)$ summarizes how i misreports when facing a given outcome. As such, $F_i(\cdot|\cdot)$ captures very generally the inherent dishonesty of person i.[1] Focusing on $F_i(\cdot|\cdot)$ is generally not a practical way of conceptualizing differences in dishonesty however - it is a high-dimensional object and is generally not possible to measure from observed behavior in a probabilistic lying game. Accordingly, the key part of the analysis below deals with how to define and construct a scalar measure of dishonesty, D_i, that summarizes the information in F_i.

2.3 Maintained assumptions on the lying game and behavior

The analysis further below will rely on some maintained assumptions that I present and discuss here.

First, because w_{it} comes from a idiosyncratic random process, w_{it} will be i.i.d. across individuals and rounds, and independent of both the individual propensity for dishonesty and the other variables in the analysis:

Assumption 2. *The outcomes* w_{it} *are i.i.d. across* i *and* t*. Moreover,* $w_{it} \perp (Y_i, X_i, F_i(\cdot|\cdot))$.

An important assumption that is invoked routinely in studies using probabilisty lying games is that participants never misreport so as to lower their payoff (no downward lying). I will invoke this assumption throughout this paper as well.

Assumption 3. $F_i(m|w) = 0$ *for all* $m < 0$ *and all* $w \in \Omega$

It will also be necessary to invoke assumptions on the dependence between the reported outcome in the game, r_{it}, and the other variables used in the analysis of interest, Y_i and X_i. Specifically, I will assume that for an individual with a given propensity for dishonesty, the reported outcomes are independent of the the other variables.

Assumption 4. $\{r_{it}\}_{t=1}^{K} \perp (Y_i, X_i) | F_i(\cdot|\cdot)$

The main content of this assumption is that it rules out any direct effects of the specific reports - and thus the payoffs - in the game on the other variables. This assumption may be violated if the payoffs in the game are large enough to generate income effects on other behavior that is captured by Y_i or X_i. In addition, the assumption rules out the possibility that simply observing specific outcomes observed in the game impacts later behavior as reflected in Y_i or X_i (for example that

[1] The one aspect of individual dishonesty that is not captured by F_i is differences in time dependence across individuals. Some individuals could conceivably tend to misreport more if they misreported a lot in the recent past, while others may behave the other way around and be less likely to lie if they have already told many lies. In spite of this, both types of individuals may well have the same marginal distribution of misreporting when examining a single round of the lying game.

participants whose die happens to show a lot of 1s change behavior after the game because they feel very unlucky). Note however that the assumption does *not* imply that propensity for dishonesty is unrelated to the other variables in the analysis. The individual distribution of misreporting $F_i(\cdot|\cdot)$ can have any relationship with X_i and Y_i.

Finally, I impose the following assumption on the sample of participants and their behavior in the lying game:

Assumption 5. $\left(Y_i, X_i, F_i(\cdot|\cdot), \{w_{it}, r_{it}\}_{t=1}^{K}\right)$ *is i.i.d. across* i

This assumption corresponds to assuming that the participants are a randomly drawn sample from the underlying population of interest. In addition, the assumption rules out that individuals' behavior and experiences in the probabilistic lying game influence each other.

2.4 Measuring propensity for dishonesty vs. behavior in the specific experiment

Before proceeding, it is worth emphasizing a subtle but important dichotomy in the use of probabilistic lying games. In the present paper the aim of the game is to construct a measure of each participants *general propensity for dishonesty* as reflected in the distribution fucntion $F_i(\cdot|\cdot)$. This is distinct from situations in which the aim is to examine the *specific misreporting*, m_{it}, that participants do during the K rounds of the game.

A simple example illustrates the difference well: A researcher conducts an experiment in which participants first report their gender, then participate in 5 rounds of a probabilistic lying game and finally play a dictator game where they are offered the possibility of donating some of their winnings to a charity. The researcher wishes to use this data to test two hypotheses: 1) that there are gender differences in the propensity for dishonesty among participants, 2) that individuals who have just behaved very honestly in a lying game feel less compelled to donate to a charity. Testing the former hypothesis is an example of the use of lying games considered here; the researcher wishes to test whether $F_i(\cdot|\cdot)$ is systematically different across genders. For the purpose of testing the latter hypothesis, however, the object of interest is m_{it} - specifically whether individuals misreported in the concrete lying game.

3 The binary case (coin-tossing, dice-guessing)

I start by considering the important case where the observed and reported outcome in each round of the lying game is binary. Without loss of generality, I assume that w_{it} is a Bernoulli random variable (so $\Omega = \{0, 1\}$) with a known succes probability of p^*. Examples of this include games where participants observe and report whether a coin toss came up tails ($p^* = \frac{1}{2}$), whether a die roll is 5 or higher ($p^* = \frac{1}{3}$) or whether the participant correctly anticipated the outcome of a die roll ($p^* = \frac{1}{6}$).

3.1 The cheat rate as a scalar measure of dishonesty

Under the assumptions of no downward lying, dishonest behavior is simple to characterize in the binary setting. In each round, behaving dishonestly means reporting $r_{it} = 1$ while in fact $w_{it} = 0$. One measure of the propensity dishonesty is therefore simply the likelihood of misreporting when $w_{it} = 0$. Following Barfort et al. (2019), I refer to this as the individual *cheat rate*:

Definition 1. The individual cheat rate is $\theta_i \equiv P_i(r_{it} = 1|w_{it} = 0) = F_i(1|0)$

In the binary case, the individual cheat rate is a natural measure of the propensity for dishonesty. Indeed, one can show that θ_i is the only possible measure of dishonesty in the sense that knowing θ_i completely determines an individual's distribution of misreproting behavior, $F_i(\cdot|\cdot)$.[2]

It is also straightforward to produce unbiased estimates of the individual cheat rates from data in the probabilistic lying game using a method of moments estimator. The estimate is simply a linear transformation of the share of reported successes (based on the known success probability p^*):

Definition 2. The estimated cheat rate is $\hat{\theta}_i \equiv \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} r_{it} - p^*\right)$

As is easily shown, the estimated cheat rate is an unbiased estimate of the true individual cheat rate but is of course subject to measurement error. I let $\eta_i \equiv \hat{\theta}_i - \theta_i$ denote the measurement error in $\hat{\theta}_i$. The following formula provides an instructive decomposition regarding the nature of this measurement error:[3]

$$\eta_i = \frac{1}{1-p^*} \underbrace{\left(\frac{1}{K}\sum_{t=1}^{K} w_{it} - p^*\right)}_{\text{Deviation in share of actual successes from the expected share}} + \frac{1}{1-p^*} \underbrace{\left(\frac{1}{K}\sum_{t=1}^{K} m_{it} - P_i(m_{it}=1)\right)}_{\text{Deviation in } i\text{'s share of misreports from the expected}}$$

[2]To see this note that because reports must be in $\{0, 1\}$, assumption 3 implies that $F_i(\cdot|\cdot)$ can be written out as:

$$F_i(m|w) = \begin{cases} 0 & \text{if } m = -1, w = 1 \\ 1 & \text{if } m = 0, w = 1 \\ \theta_i & \text{if } m = 1, w = 0 \\ 1 - \theta_i & \text{if } m = 0, w = 0 \end{cases}$$

[3]Note that we have $P_i(m_{it} = 1) = (1 - p^*)\theta_i \iff \theta_i = \frac{1}{1-p^*}(P_i(m_{it} = 1))$. The decomposition then follows from:

$$\begin{aligned} \hat{\theta}_i - \theta_i &= \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K}(r_{it} - m_{it}) + \frac{1}{K}\sum_{t=1}^{K} m_{it} - p^*\right) - \theta_i \\ &= \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} w_{it} + \frac{1}{K}\sum_{t=1}^{K} m_{it} - p^*\right) - \frac{1}{1-p^*}(P_i(m_{it} = 1)) \\ &= \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} w_{it} - p^*\right) + \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} m_{it} - P_i(m_{it} = 1)\right) \end{aligned}$$

As the formula shows, measurement error in the estimated cheat rate arises for two reasons: First, measurement error arises directly because of randomness in the actual outcomes; regardless of their individual propensity for dishonesty, some participants will be lucky and get more actual successes in the lying game than would be expected. This will result in an erroneously high estimated cheat rate. The reverse is true for participants who get very unlucky. Second, measurement error also arises because of random deviations between participants general tendency for dishonesty and their actual behavior acros the specific K rounds of the lying game; an individual who cheats on average half the time may not necessarily end up cheating in half of the rounds of a particular dice game.

Conveniently, the measurement error in the estimated cheat rate turns out to be classical in the sense that it is mean independent of the true cheat rate and all other variables included in the analysis:

Proposition 1. $E[\eta_i|\theta_i, Y_i, X_i] = 0$

Proof. Under Assumption 3 we have that $P_i(m_{it} = 1) = (1-p^*)\theta_i$. The measurement error formula can therefore be written as:

$$\eta_i = \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} w_{it} - p^*\right) + \frac{1}{1-p^*}\left(\frac{1}{K}\sum_{t=1}^{K} m_{it} - (1-p^*)\theta_i\right)$$

Now I take conditional expectations, conditional on $F_i(\cdot|\cdot)$,Y_i and X_i:

$$E[\eta_i|F_i(\cdot|\cdot), Y_i, X_i] = \frac{1}{1-p^*}E\left[\frac{1}{K}\sum_{t=1}^{K} w_{it} - p^*|F_i(\cdot|\cdot), Y_i, X_i\right] + \frac{1}{1-p^*}E\left[\frac{1}{K}\sum_{t=1}^{K} m_{it} - (1-p^*)\theta_i|F_i(\cdot|\cdot), Y_i, X_i\right]$$

Then I consider each of the two expectations terms separately. From Assumption 2, it follows that $E\left[w_{it}|F_i(\cdot|\cdot), Y_i, X_i\right] = E[w_{it}] = p^*$. This implies $E\left[\frac{1}{K}\sum_{t=1}^{K} w_{it} - p^*|F_i(\cdot|\cdot), Y_i, X_i\right] = 0$

Assumptions 2 and 4, imply that $E\left[m_{it}|F_i(\cdot|\cdot), Y_i, X_i\right] = E[m_{it}|F_i(\cdot|\cdot)] = (1-p^*)\theta_i$ so that also $E\left[\frac{1}{K}\sum_{t=1}^{K} m_{it} - (1-p^*)\theta_i|F_i(\cdot|\cdot), Y_i, X_i\right] = 0$

It follows that $E[\eta_i|F_i(\cdot|\cdot), Y_i, X_i] = 0$. A (trivial) application of the law of iterated expectations completes the argument. □

As I expand on in the next section, the fact that the measurement error in estimated cheat rates is classical means that it has the usual, well-understood consequences when estimated cheat rates are used in linear regression frameworks. Before proceeding to discuss this, however, it is worth making the following remarks about the estimated cheat rate:

1. The description above assumes that data from all K rounds of the probabilistic lying game are used to construct the estimated cheat rate. All the arguments above (and everything that follows in the next subsection) go through if data on only some subset of these rounds are

used when constructing the estimated cheat rate. This is useful for robustness checks and for addressing potential concerns that the maintained assumptions do not apply across all rounds. For example, researchers may in practice worry that there is some learning occuring during the first rounds and/or fatigue in later rounds that violates one or more of the assumptions above. If so, these rounds can simply be excluded.

2. The arguments above and in the next subsection go through also in the case where $K = 1$. The approaches considered here can thus be applied using data from only a single round of a lying game.

3. Becase the estimated cheat rate is an unbiased estimator, the sample mean of the estimated cheat rate is an unbiased estimate of the mean cheat rate in the underlying population. This provides a straightforward way to characterize the average level of dishonesty.

3.2 Using the estimated cheat rate in linear regression frameworks

I now consider estimation of the two linear regressions of interest using the cheat rate as the scalar defintion of dishonesty and using the estimated cheat rate as a the empirical measure. That is I set, $D_i = \theta_i$ and consider estimating the regressions of interest using $\hat{\theta}_i$ as a regressor.

Plugging in for θ_i in equation 2 I get:

$$\hat{\theta}_i = \beta^{LHS} Y_i + X_i' \pi^{LHS} + e_i^{LHS} \quad , \quad e_i^{LHS} \equiv \varepsilon_i^{LHS} + \eta_i \tag{3}$$

From Proposition 1, it follows that the composite error term in this regression satisfies $E\left[e_i^{LHS} | Y_i, X_i\right] = 0$. Accordingly, OLS estimation of this regression will recover consistent estimates of β^{LHS} as usual.

Plugging in for θ_i in equation 1 I get:

$$Y_i = \beta^{RHS} \hat{\theta}_i + X_i' \pi^{RHS} + e_i^{RHS} \quad , \quad e_i^{RHS} \equiv \varepsilon_i^{RHS} - \beta \eta_i \tag{4}$$

Unfortunately, this composite error term is not mean independent of regressors, implying that OLS will in general be inconsistent because of the measurement error in D_i. Indeed the standard attenuation bias formula for univariate regression applies.[4] The next sections presents a simple instrumental variables approach to addressing the bias arising from measurement error here.

[4]Consider the case with no control variables in which $X_i = 1$. In this case, the OLS estimate of β^{RHS} has the probability limit $\frac{Cov(Y_i, \hat{D_i})}{Var(\hat{D_i})}$. Now following a standard argument:

$$\begin{aligned} \frac{Cov(Y_i, \hat{\theta_i})}{Var(\hat{\theta_i})} &= \frac{Cov(\beta^{RHS}\theta_i + \varepsilon_i^{RHS}, \theta_i + \eta_i)}{Var(\theta_i + \eta_i)} \\ &= \beta^{RHS} \frac{Cov(\theta_i, \theta_i)}{Var(\theta_i + \eta_i)} = \beta^{RHS} \frac{Var(\theta_i)}{Var(\theta_i) + Var(\eta_i)} \end{aligned}$$

3.3 Addressing measurement error using a split sample instrument

I now present a simple approach to addressing the measurement error problem in equation 4, specifically using an instrument based on splitting the data from the probabilistic lying game. The approach only requires an additional assumption on the time dependence on cheating behavior. Specifically, I assume that for an individual with some given level of dishonesty, any dependence between misreporting behavior over time dies out after $\kappa \geq 0$ rounds of the lying game, where $\kappa < K - 2$. I also assume that the same holds for the dependence between past observed outcomes and misreporting behavior, that is, outcomes observed more than κ rounds ago can not influence your current decision about whether to misreport:

Assumption 6. *For all t and t' such that $|t - t'| > \kappa$, the following holds: $m_{it} \perp m_{it'}|F_i(\cdot|\cdot)$ and $m_{it} \perp w_{it'}|F_i(\cdot|\cdot)$*

Now let $\tilde{\theta_i^1}$ be the estimated cheat rate for person i when using data only onthe first $[(K-\kappa)/2]$ rounds and let $\tilde{\theta_i^2}$ denote the corresponding estimated cheat rate using only data from the last $[(K-\kappa)/2]$ rounds. Here $[\cdot]$ is the integer part function and simply deals with the possibility that $(K-\kappa)$ is odd. Also let $\tilde{\eta_i^1}$ and $\tilde{\eta_i^2}$ denote the corresponding measurement error. Formally:

Definition 3.

$$\tilde{\theta_i^1} \equiv \frac{1}{1-p^*}\left(\frac{1}{[(K-\kappa)/2]}\sum_{t=1}^{[(K-\kappa)/2]} r_{it} - p^*\right)$$
$$\tilde{\theta_i^2} \equiv \frac{1}{1-p^*}\left(\frac{1}{[(K-\kappa)/2]}\sum_{t=K-[(K-\kappa)/2]+1}^{K} r_{it} - p^*\right)$$
$$\tilde{\eta_i^1} \equiv \tilde{\theta_i^1} - \theta_i$$
$$\tilde{\eta_i^2} \equiv \tilde{\theta_i^2} - \theta_i$$

Under the additional assumption 6, it is straightforward to show that the measurement error in the two different estimated cheat rates turns out to be uncorreleated, meaning that any correlation between them purely reflects variation in actual cheat rates:

Proposition 2.

$$Cov(\tilde{\eta_i^1}, \tilde{\eta_i^2}) = Cov(\tilde{\eta_i^1}, \tilde{\theta_i^2}) = 0$$
$$Cov(\tilde{\theta_i^1}, \tilde{\theta_i^2}) = Var(\theta_i)$$

This in turn reveals a straightforward way to consistently estimate β^{RHS} via 2SLS. One of the two estimated cheat rates is used as the measure of dishonesty in the equation of interest and is instrumented using the other estimated cheat rate. Specifically, this involves the following standard two-equation model:

$$
\begin{aligned}
Y_i &= \beta^{RHS}\tilde{\theta_i^1} + X_i'\pi^{RHS} + u_i^{RHS} \quad , \quad u_i^{RHS} \equiv \varepsilon_i^{RHS} - \beta\tilde{\eta_i^1} \\
\tilde{\theta_i^1} &= \rho_0 + \rho_1\tilde{\theta_i^2} + \xi_i \quad , \quad Cov(\xi_i, \tilde{\theta_i^2}) = 0
\end{aligned}
$$

Proposition 2 implies that $Cov(\tilde{\eta_i^1}, \tilde{\theta_i^2}) = 0$ and that $\rho_1 = \frac{Var(\theta_i)}{Var(\theta_i)+Var(\tilde{\eta_i^2})} > 0$ such that 2SLS estimation of β^{RHS} will provide consistent estimates. Thus the simple split-sample instrument addresses the measurment error problem.

It is worth noting the following here:

1. The exposition above assumes that the split sample instrument involves two cheat rate estimates based on the same number of rounds either at the very beginning or the very end of the probabilistic dice game. All the arguments above go through however if the two cheat rates are based on other subsets of rounds, as long as all the rounds involved in the two different estimates are at least κ apart (to ensure independence). If $K - \kappa$ is odd, it is natural to include an extra round in one of the two estimates for efficiency reasons. Specific choices for the subsets may also be based on the practical concerns about learning or fatigue over rounds that were discussed previously.

2. While the 2SLS approach outlined above is transparent and simple, the setup here also lends itself naturally to an application of the ORIV estimator of Gillen et al. (2019) or a more general GMM approach. These alternative approaches may be attractive on efficiency grounds.

3. The approach above is feasiable as long as the number of rounds in the lying game satisfies $K - \kappa \geq 2$. Although the estimated cheat rates can be constructed when $K = 1$ the possibility of addressing biases from measurement error highlights an important advantage of using probabilistic lying games with more rounds.

4. Because $Cov(\tilde{\theta_i^1}, \tilde{\theta_i^2}) = Var(\theta_i)$, the approach above also reveals a natural estimator for characterizing the variance of dishonesty in the population: simply compute the empirical covariance between the two different estimated cheat rates. This in turn also makes it possible to estimate the variance of the measurement error.

3.4 Alternative measures of dishonesty

As discussed above, the (estimated) cheat rate is a natural way to define and measure dishonesty. Other meaning definitions and measures are possible however and have been used in prior work.

The cheat rate measures how often an individual misreports when the true outcome is $w_{it} = 0$ so that (advantageous) cheating is in fact possible. A different way to measure and define dishonesty is to instead focus on how often the individual misreports overall, while ignoring the fact that in rounds where the true outcome is $w_{it} = 1$ the individual is unable to cheat. To distinguish this definition of dishonesty from the cheat rate, I refer to it as the overall *misreporting rate, γ_i,:*

Definition 4. The misreporting rate is $\gamma_i \equiv P_i(m_{it} > 0) = (1 - p^*)F_i(1|0)$

Although this definition is conceptually different from the cheat rate, it has a very simple linear relationship to it. As is clear from the defintions, the misreporting rate is simply the cheating rate scaled down by $1 - p^*$. The intuition behind this is simple: The misreporting rate measure how often misreporting occurs overall. A fraction $1 - p^*$ of the time $w_{it} = 1$ and misreporting can not occur regardless of how often the individual is inclined to cheat.

As with the cheat rate, a simple method of moments estimator provides unbiased estimates of the misreporting rate. This turns out to simply be the share of rounds where the individual reports success minus the expected share under full honesty:

Definition 5. The estimated misreporting rate is $\hat{\gamma}_i \equiv \frac{1}{K}\sum_{t=1}^{K} r_{it} - p^*$

Importantly, there is the same linear relationship between the estimated misreporting rate and the estimated cheat rate, $\hat{\gamma}_i = (1-p^*)\hat{\theta}_i$. It is straightforward to see that all the previous conclusions regarding the cheat rate and estimated cheat rate therefore carry over to the misreporting rate and the estimated misreporting rate. The consequences of using the (estimated) misreporting rate instead of the (estimated) cheat rate in the linear regression frameworks are also straightforward: Using the (estimated) misreporting rate will simply scale down the parameter of interest and all estimates of it by $(1 - p^*)$. Although there may be expositional reasons to prefer one of these measures over the other, the choice between the two approaches is thus less important. Results from studies using one approach can be easily compared to studies using the other approach by simply rescaling them.

Finally, previous work using probabilistic lying games have employed a number different intuitive measures of dishonesty without relying on a formal framework and definition. Conveniently, most of these measures are easy to relate to the formal framework presented here. Specifically, some studies use simply the win share across the K rounds of the lying game as their measure of dishonesty. As is clear from above, this measure differs from the estimated misreporting rate only by the constant p^*, accordingly results from these studies are straightforward to reinterpret in light of either of the formal definitions of dishonesty provided here. The same is true for studies using the total number of successes as the measure of dishonesty.

Other studies use the total winnings in the dice game as their measure of dishonesty. Assuming that winnings is a linear function of the number of successes (as is typically the case), it is easy to see that total winnings will simply be a known linear (affine) transformation of the estimated cheat rate and/or the estimated misreporting rate. Reinterpreting these results in terms of the formal

defintion of dishonesty provided here is therefore also straightforward by simply rescaling estimates appropriately.

4 The non-binary case (reporting dice rolls)

Next, I consider the case where the observed and reported outcome in each round of the lying game is not binary. The prime example here is the case where individuals simply report a dice roll. To keep things general, however, I will simply let $G(\cdot)$ denote the cumulative distribution function for w_{it}. Moreover, I will let w^{max} denote the largest possible outcome ($w^{max} \equiv max_{w \in \Omega}\, w$) and let $\bar{w}$ denote the expected value of the actual otucome ($\bar{w} \equiv E[w_{it}]$). These will of course be known quantities to the researcher. If participants are reporting the outcome of a single die, we have $w^{max} = 6$ and $\bar{w} = 3.5$.

4.1 The average misreport as a scalar measure of dishonesty

As opposed to the binary case, where there is a single way to misreport, dishonesty can in general take many forms in the non-binary case. For the case of reporting dice rolls example, some individuals may choose to overreport all their rolls slightly, others may choose to report 6 every other time, others may again decide to simply misreport 6 whenever they in fact get a one, etc. There is no unique way of collapsing these many types of dishonesty into a meaningful scalar concept.

A natural candidate for a scalar defintion of dishonesty however is simply the amount that an individual overreports on average. I will refer to this as the *average misreport* and denote it by μ_i:

Definition 6. The average misreport for person i is $\mu_i \equiv E_i\left[m_i\right]$

Clearly this scalar definition of dishonesty is not the only possible one in this setting. In particular, I note that this scalar definition fails to distinguish between the (extensive margin) decision of whether to be dishonest at all and the (intensive margin) decision about how dishonest to be in a given situation. I return to this in Section 4.4. As will be clear below, however, the average misreport turns out to have some useful properties.

As was the case for the cheat rate defined in Section 3.1, it is straightforward to obtain unbiased estimates of the average misreport using a method of moments estimator. I refer to this as the *estimated average misreport, $\hat{\mu}_i$:*

Definition 7. The estimated average misreport for person i is $\hat{\mu} \equiv \frac{1}{K}\sum_{t=1}^{K} r_{it} - \bar{w}$

Again, although it is unbiased, the estimated average misreport will suffer from measurement error. I will denote this ν_i. such that formally, $\nu_i \equiv \hat{\mu_i} - \mu_i$. Simple derivations show that a decomposition holds for the measurement error, analogous to Section 3.1:

$$\nu_i = \underbrace{\left(\frac{1}{K}\sum_{t=1}^{K} m_{it} - \bar{w}\right)}_{\text{Deviations of the average actual outcome from the expected}} + \underbrace{\left(\frac{1}{K}\sum m_{it} - \mu_i\right)}_{\text{Deviations of the average misreport from the expected}}$$

The interpretation and intuition here is as in Section 3.1. An proof analagous to the one in Section 3.1 also shows that the measurement error is classical in this case:

Proposition 3. $E[\nu_i|\theta_i, Y_i, X_i] = 0$

Finally, the remarks from the end of Section 3.1 also apply here: The arguments above go through when using only a subset of rounds and when $K = 1$. Moreover, the sample average of $\hat{\mu_i}$ is an unbiased estimate of average dishonesty as defined in Definition 6.

4.2 Using the estimated average misreport in linear regression frameworks

Using the (estimated) average misreport as the measure of dishonesty, the issues and solutions for the linear regression framework turn out to be trivial extensions of the binary case. Setting $D_i = \mu_i$ in equation 2 and substituting in for μ_i yields the following:

$$\hat{\mu_i} = \beta^{LHS} Y_i + X_i' \pi^{LHS} + e_i^{LHS} \quad , \quad e_i^{LHS} \equiv \varepsilon_i^{LHS} + \nu_i \tag{5}$$

As before, the composite error term satisfies $E\left[e_i^{LHS}|Y_i, X_i\right] = 0$ so OLS will provide consistent estimates

Doing the same in equation 1, however, again reveals that OLS will be (asumptotically) biased when dishonesty is on the right hand side of the regression (with the usual formula applying for the univariate case). The bias can be addressed using the split-sample instrument approach. We can again define two separate estimates of the average misreport using data from different subset of rounds:

Definition 8.

$$\tilde{\mu_i^1} \equiv \frac{1}{[(K-\kappa)/2]} \sum_{t=1}^{[(K-\kappa)/2]} r_{it} - \bar{w}$$

$$\tilde{\mu_i^2} \equiv \frac{1}{[(K-\kappa)/2]} \sum_{t=K-[(K-\kappa)/2]+1}^{K} r_{it} - \bar{w}$$

$$\tilde{\nu_i^1} \equiv \tilde{\mu_i} - \mu_i$$

$$\tilde{\nu_i^2} \equiv \tilde{\mu_i^2} - \mu_i$$

Invoking assumption 6, we then again have that the measurement error in the two estimates are uncorrelated:

Proposition 4.

$$Cov(\tilde{\eta_i^1}, \tilde{\eta_i^2}) = Cov(\tilde{\eta_i^1}, \tilde{\mu 2}) = 0$$
$$Cov(\tilde{\mu_i}, \tilde{\mu_i^2}) = Var(\mu_i)$$

This in turn implies that 2SLS estimation of the following two-equation model will yield consistent estimates:

$$Y_i = \beta^{RHS}\tilde{\theta_i^1} + X_i'\pi^{RHS} + u_i^{RHS} \quad , \quad u_i^{RHS} \equiv \varepsilon_i^{RHS} - \beta\tilde{\nu_i}$$
$$\tilde{\theta_i^1} = \rho_0 + \rho_1\tilde{\theta_i^2} + \xi_i \quad , \quad Cov(\xi_i, \tilde{\gamma_i^2}) = 0$$

The previous remarks about choosing different subsets of rounds, using a smaller K, employing the ORIV estimator of Gillen et al. (2019) and about estimating $Var(\mu_i)$ also apply (see Section 3.3).

4.3 Alternative scalar measures of dishonesty in the non-binary case

In the context of binary lying games, Section 3.4 emphasized the distinction between measures of dishonesty that focus on how often dishonesty occurs when it is possibly (e.g. the cheat rate) and measures that focus on how often dishonesty occurs overall (e.g. the misreporting rate). A version of this distinction can be applied in the non-binary case as well. The average misreport, μ_i, falls in the second category. It measures on average how much a person misreports by without accounting for the fact that when $w_{it} = w^{max}$ it is not possible to misreport (upwards).

It is possible to introduce a scalar definition of dishonesty that instead accounts for this also in the non-binary case. I define the *average misreport when possible,* α_i, as the mean misreport for person i when the actual outcome is not $w_{it} = w^{max}$:

Definition 9. The *average misreport when possible* is $\alpha_i \equiv E_i\left[m_i | w_{it} < w^{max}\right]$

As for the binary case, there turns out to be a convenient, simple linear relationship between this scalar measure of dishonesty and the one considered in previous sections (the average misreport, μ_i). Specifically we have:[5]

[5]To see this note that:

$$\begin{aligned}\mu_i = E\left[m_{it}\right] &= G(w^{max}) \times 0 + (1 - G(w^{max}))\, E_i\left[m_i | w_{it} < w^{max}\right] \\ &= (1 - G(w^{max}))\, E_i\left[m_i | w_{it} < w^{max}\right] = (1 - G(w^{max})\mu_i\end{aligned}$$

$$\mu_i = (1 - G(w^{max})\alpha_i$$

Completely analagously to the binary case, we can also define an *estimated average misreport when possible* based on an unbiased method of moments estimator, which turns out also be a linear transformation the estimated average misreport. All the previous results thus again carry over and there is a straightforward way to compare empirical results using either of these defintions of dishonesty simply by rescaling estimates appropriately.

Finally, we can relate some standard intuitive measures to the estimated average misreport. The average reported outcomes or the sum of the reported outcomes over the K rounds are simply linear (affine) transformation of the estimated average misreport, which again makes it easy to reinterpret estimates using these measures simply via rescaling. Assuming that winnings are linear in the reported rolls, the same is true for average or total winnings.

4.4 The extensive vs. intensive margin of dishonesty

As hinted at previously, an important feature of the non-binary lying game is that it introduces both an intensive and an extensive margin of dishonesty. In addition to choosing whether to misreport on a given roll, the individual must choose how much to misreport by. Using the average misreport as a scalar measure of dishonesty collapses behavior in both of these dimensions into one.

To see this more formally, we can start by defining the conditional average misreport, where the conditioning is on misreporting by at least some positive amount. This is a natural measure of intensive margin dishonesty:

Definition 10. For an individual i where $P(m_i > 0) > 0$, the conditional average misreport is $\delta_i \equiv E_i\left[m_i | m_i > 0\right]$

With this definition in place and invoking also the definition of the misreporting rate from Section 3.4, we can decompose an individual's average misreport into an intensive and extensive margin component as follows:

$$\begin{aligned}\mu_i = E_i\left[m_i\right] &= \underbrace{P_i(m_i > 0)}_{\text{Likelihood of misreporting}} \times \underbrace{E_i\left[m_i | m_i > 0\right]}_{\text{Average misreport when misreporting}} \\ &= \gamma_i \times \delta_i\end{aligned}$$

The first factor, the misreporting rate, represents the the extensive margin decision about how often an individual actually misreports by any amount. The second factor, the conditional average misreport, represents the intensive margin decision about how much to misreport.

A natural question then is whether there exists empirical measures of dishonesty based on non-binary lying games that reliable separates out the extensive margin dishonesty, γ_i, from the

instensive margin dishonesty, δ_i. The general answer turns out to be no unfortunately. Using data from a standard non-binary proabilistic lying game, there is no way to distinguish individuals who often misreport a small amount (high γ_i, low δ_i) from individuals who rarely misreport but misreports by a large amount when they do (low γ_i, high δ_i) .[6] In the next section, I return to consider the case where actual outcomes are potentially observed by the researcher alongside participants' reports, which makes it possible to make progress on this issue.

Note that, as shown in Section 3.4, it is straightforward to construct an unbiased estimate of γ_i in the context of a binary lying game (i.e. the estimated misreporting rate, $\hat{\gamma_i}$), while no such unbiased estimator is available in the context of a non-binary lying game. This highlights a key difference between binary and non-binary lying games. Binary lying games make it possible to measure and study extensive margin dishonesty in isolation. In contrast, non-binary lying games makes it possible to measure and study (only) measures of dishonesty that combine intensive and extensive margin dishonesty.

4.5 Observing the actual outcomes

I finish the discussion of the non-binary cases by considering the special case where researchers may for some reason directly observe the actual outcomes, w_{it}, alongside the individuals' reports, r_{it}. This implies that the researcher can directly infer misreporting with certainty. Observing lying or misreporting with certainty is commonplace in some other types of experimental lying or cheating games (see Jacobsen et al. (2018) for a review). It is typically seen as strength of the probabilistic lying games design that participants are completely certain that they can never be caught misreporting. As illustrated by the seminal implementation in Kroell and Rustagi (2021), however, it may be possible to preserve this feature of probabilistic lying games and still observe actual misreports by using an electronic (bluetooth) die that electroncially reports outcomes without participants knowing. As it turns out, observing the actual outcomes offers some particular advantages in the non-binary case thus I focus on this case here.

In the context of the econometric framework used here, if the researcher observes both reports, r_{it}, and actual outcomes, w_{it}, the individual misreport can be inferred directly as well, $m_{it} = r_{it} - w_{it}$. As is immediately clear, this makes it possible to construct an alternative estimator for the average misreport simply using the mean misreport in the data for each person as:

$$\hat{\mu_i^{obs}} \equiv \frac{1}{K} \sum_{t=1}^{K} m_{it}$$

[6] Consider the following example in the context of reporting the outcome of a fair die: Individual 1 misreports 1 higher than the actual roll whenever he rolls 1-5. Individual 2 misreports a 6 whenever he rolls a 1 but never misreports otherwise. Individual 1 misreports much more often (5 times more often) than individual 2. When individual 2 does misreport however he misreports by a much larger amount. The full distribution of reports - as well as the average misreport - for these two individuals will be identical however. The researcher therefore has no way of distinguishing them with the available data.

As is easily seen, this alternative estimator is also unbiased. Less obvious perhaps, it formally still suffers from measurement error. In particular we have:

$$\hat{\mu_i^{obs}} - \mu_i = \underbrace{\left(\frac{1}{K}\sum m_{it} - \mu_i\right)}$$

Comparing this expression for the measurement error in $\hat{\mu_i^{obs}}$ to the one shown previously for $\hat{\mu_i}$ we see that using the observed misreports gets rid of the measurement error stemming simply from deviations in the actual outcomes w_{it} from their average (i.e. getting lucky/unlucky in the actual outcomes). The second source of measurement error still remains, however. An individual who in general tends to misreport by 1 will not necessarily misreport by exactly that amount on average across the K specific rounds of the lying game. This introduces measurement error even though actual misreporting is observed.

I note that the measurement error in $\hat{\mu_i^{obs}}$ will obviously be smaller than in $\hat{\mu_i}$. It may even be small enough that a researcher is willing to ignore it, as is often done in similar situations. Formally, however, $\hat{\mu_i^{obs}}$ will be subject to measurement error. In any case, the split-sample instrument is straightforward to adapt here to address measurement error.

Besides the reduction in measurement error, a more interesting implication of observing the actual outcomes is that it opens the possibility of systematically separating intensive and extensive margin dishonesty. Specifically, in the context of a non-binary lying game with observed outcomes, it is possible to separately construct an unbiased estimator of the extensive margin dishonesty as measured by the misreporting rate. Letting $I(\cdot)$ denote the indicator function, this can be done simply as:

$$\hat{\gamma_i^{obs}} \equiv \frac{1}{K}\sum_{t=1}^{K} I(m_{it} > 0)$$

A natural approach to estimating intensive margin dishonesty as measured by the conditional average misreport also exists:

$$\hat{\delta_i^{obs}} \equiv \frac{1}{|\mathcal{M}_i|}\sum_{t\in\mathcal{M}_i} m_{it} \quad , \quad \mathcal{M}_i = \{t : w_{it} > 0\}$$

Implementing this latter estimator of course requires dealing with the technical issue that δ_i is not naturally defined for individuals who never cheat and correspondingly that δ_i^{obs} as written is not defined for individuals who is never observed cheating in the data.

Given how rare it is to have data on observed outcomes, I do not pursue this approach further here but simply note that it has great promise for answering questions about whether (and how)

innate dishonesty along the extensive margin may differ from innate dishonesty along the extensive margin.

5 Conclusion and practical recommendations

This paper presents a simple econometric framework that describes probabilistic lying games and relates behavior in these games to a general defintion of dishonesty. The framework is used to discuss useful ways to define and measure inherent dishonesty using these games, as well as how to include the resulting measures in linear regression frameworks and deal with measurement error.

Under a set of standard assumptions, the results in the paper sugges the following practical advice for future work:

- If the research question at hand can be answered by focusing on extensive margin dishonesty (whether to lie), binary probabilisticy lying games offer a unique way to measure individual dishonesty, as defined by the cheat rate (see Definition 1).
- If the research question at requires that the analysis captures also intensive margin dishonesty (how much to lie), non-binary probabilistic lying games make it possible to measures individual dishonesty, using the concept of the average misreport (see Definition 6).
- Unbiased estimates of the cheat rate or the average misreport are easily computed from observed behavior in the lying games, even if each participant only participates in one round of the game (see Defintions 2 and 7)
- The estimated cheat rates or average misreports are subject to measurement error. The measurement error is classical, however, meaning that linear regression estimates remain consistent if dishonesty is used as the outcome variable.
- If dishonesty is used as the independent variable in a linear regression, measurement error generates attenuation bias. If each participant participates in more than one round of the lying game however, a simple split-sample instrument can be used to remove this bias if an additional restriction is imposed on the time dependence of dishonest behavior (See Sections 3.3 and 4.2)
- Because most measures used in past work are linear transformation of either the estimated cheat rate or the estimated average report, comparisons with previous work are straightforward by simply rescaling linear regression estimates (see Sections 3.4 and 4.3).

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Barfort, Sebastian, Nikolaj A. Harmon, Frederik Hjorth, and Asmus Leth Olsen, "Sustaining Honesty in Public Service: The Role of Selection," *American Economic Journal: Economic Policy*, November 2019, *11* (4), 96–123.

Cohn, Alain and Michel André Maréchal, "Laboratory Measure of Cheating Predicts School Misconduct," *The Economic Journal*, 03 2018, *128* (615), 2743–2754.

Fischbacher, Urs and Franziska Föllmi-Heusi, "LIES IN DISGUISE—AN EXPERIMENTAL STUDY ON CHEATING," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Gächter, Simon and Jonathan F. Schulz, "Intrinsic honesty and the prevalence of rule violations across societies," *Nature*, 2016, *531* (7595), 496–499.

Gans-Morse, Jordan, "Self-Selection into Corrupt Judiciaries," *Journal of Law, Economics and Organization*, Forthcoming.

_ , Alexander Kalgin, Andrei Klimenko, Dmitriy Vorobyev, and Andrei Yakovlev, "Self-Selection into Public Service When Corruption is Widespread: The Anomalous Russian Case," *Comparative Political Studies*, 2021, *54* (6), 1086–1128.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*, 2019, *127* (4).

Hanna, Rema and Shing-Yi Wang, "Dishonesty and Selection into Public Service: Evidence from India," *American Economic Journal: Economic Policy*, August 2017, *9* (3), 262–90.

Jacobsen, Catrine, Toke Reinholt Fosgaard, and David Pascual-Ezama, "WHY DO WE LIE? A PRACTICAL GUIDE TO THE DISHONESTY LITERATURE," *Journal of Economic Surveys*, 2018, *32* (2), 357–387.

Kroell, Markus and Devesh Rustagi, "Measuring Honesty and Explaining Adulteration in Naturally Occuring Markets," *Working Paper*, 2021.

Moshagen, Morten and Benjamin E. Hilbig, "The statistical analysis of cheating paradigms," *Behavior Research Methods*, 2017, *49* (2), 724–732.

Potters, Jan and Jan Stoop, "Do cheaters in the lab also cheat in the field?," *European Economic Review*, 2016, *87*, 26–33.