Behav Res (2017) 49:724–732
DOI 10.3758/s13428-016-0729-x

The statistical analysis of cheating paradigms

Morten Moshagen[1] **· Benjamin E. Hilbig**[2,3]

Published online: 11 April 2016

Abstract One type of paradigm commonly used in studies on unethical behavior implements a lottery, relying on a randomization device to determine winnings while ensuring that the randomized outcome is only known to participants. Thereby, participants have the incentive and opportunity to cheat by anonymously claiming to have won. Data obtained in such a way are often analyzed using the observed "win" responses as a proxy for actual dishonesty. However, because the observed "win" response is contaminated by honest respondents who actually won, such an approach only allows for inferring dishonesty indirectly and leads to substantially underestimated effects. As a remedy, we outline approaches to estimate correlations between dishonesty and other variables, as well as to predict dishonesty in a modified logistic regression model. Using both simulated and empirical data, we demonstrate the superiority and relevance of the suggested methods.

Keywords Behavioral ethics · Cheating · Correlation · Dishonesty · Logistic regression · Inference

Dishonest, deceitful, and fraudulent behaviors incur noteworthy challenges for societies (Mazar & Ariely, 2006; Rosenbaum, Billinger, & Stieglitz, 2014)—such as billions lost annually to individual income tax evasion (Mazur & Plumley, 2007). Whereas the study of unethical behavior was long characterized by a normative focus, recently an upsurge of interest has coincided with a more descriptive view. This approach—spanning several subfields of economics, management, psychology, and other social sciences—is now commonly referred to as *behavioral ethics* (Bazerman & Gino, 2012). Its goal is to understand the determinants, boundary conditions, and underlying processes of unethical behavior (especially cheating and dishonesty), and numerous recent studies have shed light on such issues (e.g., Bryan, Adams, & Monin, 2013; Caruso & Gino, 2011; Gino & Margolis, 2011; Hershfield, Cohen, & Thompson, 2012; Mazar, Amir, & Ariely, 2008; Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009; Peer, Acquisti, & Shalvi, 2014; Schurr, Ritov, Kareev, & Avrahami, 2012; Schweitzer, Ordóñez, & Douma, 2004; Shalvi, Eldar, & Bereby-Meyer, 2012; Shalvi, Handgraaf, & De Dreu, 2011; Vohs & Schooler, 2008; Weisel & Shalvi, 2015; Zhong, Bohns, & Gino, 2010; Zimerman, Shalvi, & Bereby-Meyer, 2014).

One particularly advantageous type of paradigm commonly used in studies on unethical behavior relies on a probabilistic link between observed response behavior and cheating or dishonesty (e.g., Abeler, Becker, & Falk, 2014; Bryan et al., 2013; Conrads, Irlenbusch, Rilke, & Walkowitz, 2013; Fischbacher & Heusi, 2008; Hilbig & Hessler, 2013; Lewis et al., 2012; Nogami & Yoshida, 2013; Peer et al., 2014; Shalvi et al., 2012; Shalvi et al., 2011): Specifically, participants take part in a lottery using some sort of randomization device (e.g., rolling a die, tossing a coin, etc.) to determine whether they are eligible for a reward (typically monetary gains). Importantly, the setup ensures that the true outcome of the randomization process is only known to the participant, but not to the investigators (e.g., by rolling the die in secret). Thereby, participants have both the opportunity and incentive

✉ Morten Moshagen
moshagen@uni-kassel.de

[1] Research Methods, Institute of Psychology, University of Kassel, Holländische Straße 36-38, 34127 Kassel, Germany

[2] University of Koblenz-Landau, Landau, Germany

[3] Max-Planck-Institute for Research on Collective Goods, Bonn, Germany

to cheat by claiming to have obtained the target outcome, despite actually having obtained a different outcome. Thus, the actual honesty of each single participant's response remains unknown, as the only observed outcomes from using such a paradigm are participants' "win" or "not win" responses. For example, participants in a study by Hilbig and Zettler (2015, Exp. 2) were instructed to roll a fair die in secret. They were promised a monetary reward if they *reported* a particular target number. Although the probability of obtaining a legitimate win thus was merely p = 1/6, almost half of the participants (45 %) reported the target number and thus allegedly won. Given the anonymity of the experimental situation, no statement is possible regarding whether a particular individual might have cheated. However, the proportion of dishonest individuals can be inferred from these observed "win" or "not win" responses at the aggregate level, provided that the probability distribution of the randomization device is known and held constant (Moshagen, Hilbig, Erdfelder, & Moritz, 2014). To this end, it is generally assumed that (1) dishonest respondents always claim to have won regardless of the outcome, whereas honest respondents only report to have won if actually having (rightfully) obtained the target outcome, and (2) participants do not lie to their disadvantage by denying to have won despite actually having obtained the target outcome (but see Fischbacher & Utikal, 2011). Given these assumptions, it is straightforward to estimate different proportions at group level from the observed proportion of alleged wins (see below for details).

However, the actual status of each individual who claimed a win necessarily remains unknown. Since the observed proportion of alleged wins suffers from contamination by honest respondents who actually did win, this precludes the immediate interpretation of the observed "win" responses as an indicator of cheating or dishonesty. Nevertheless, studies employing such a paradigm consistently rely on the observed "win" responses as the dependent variable in a regression framework. Although nothing is inherently wrong with this approach, it allows for only indirect conclusions about dishonesty, because the probability of winning, not the probability of dishonesty, is predicted. Of course, to the extent that the observed or predicted probability of winning departs from the baseline probability, dishonesty is a prime candidate to account for this difference—given the assumptions stated above. Nonetheless, on a theoretical level, it would be preferable to predict dishonesty, rather than alleged wins. Moreover, ignoring that the observed "win" response is contaminated by honest respondents leads to substantially underestimated relationships between dishonesty and other variables (such as the covariates reflecting different experimental conditions). Although such underestimations lead to conservative conclusions, it is obviously preferable to apply analytic methods that would allow for an undistorted assessment.

In light of these deficiencies, a superior approach would be to predict dishonesty, rather than alleged wins. In the remainder of this article, we outline methods to estimate the proportion of dishonest individuals, to compute correlations between dishonesty and other variables, and to predict dishonesty using an adapted logistic regression framework. We then report on a simulation study demonstrating that the proposed methods outperform traditional approaches. Finally, we also illustrate the methods by reanalyzing published studies and provide brief code examples to perform the described analyses in R (see the Appendix).

Estimating the prevalence of dishonesty

Say we ask 100 participants to toss a fair coin once (in secret) and to report only the outcome, promising an incentive for all who report tails. Since the coin is fair, we know that half of all 100 tosses will most likely have turned up tails—that is, the baseline probability of winning is p = .50. Suppose that 75 respondents claimed to have won, so that the observed proportion of alleged wins would be q = .75. The probability of observing a "win" response is a function of the proportions of honest and dishonest individuals as well as of the baseline probability p, as dictated by the chosen randomization device. Dishonest respondents are assumed to always report a win (regardless of the actual outcome), whereas honest respondents will only report a win if they have actually obtained the target outcome. More formally,

$$p(\text{“win”}) = d + (1-d)\cdot p, \tag{1}$$

where d denotes the proportion of dishonest individuals (so that $1 - d$ is the proportion of honest respondents). Solving for d yields

$$d = \frac{p(\text{“win”})-p}{(1-p)}, \tag{2}$$

so an estimate for the proportion of dishonest individuals can be obtained by replacing p("win") with the observed sample proportion of reported wins (q). In the example, the estimate of the proportion of dishonest individuals is $\hat{d} = (075-.50)\ /\ (1-.50) = .50$. The associated standard error can be obtained from

$$SE^2 = \frac{q\cdot(1-q)}{N-(1-p)^2}, \tag{3}$$

and the G^2 statistic for the test of whether d is equal to zero is

$$G^2 = 2\cdot N\cdot\left[q\cdot\ln\left(\frac{N\cdot q}{N\cdot p}\right) + (1-q)\cdot\ln\left(\frac{N\cdot(1-q)}{N\cdot(1-p)}\right)\right], \tag{4}$$

which is referred to a chi-square distribution with one degree of freedom (Read & Cressie, 1988).[1] Thus, in the present example, SE = .09, 95 % CI = .33–.67, and $G^2(1) = 26.16$, $p < .01$.

Note that d refers to the proportion of respondents who would cheat if required (i.e., if they did not actually win), defined contingent on the observation that winning must be equally likely for honest and dishonest individuals. Thus, the rate of actual cheating behavior c (the proportion of factually untruthful responses) immediately follows from the proportion of dishonest individuals, $c = (1-p) \cdot d$. Because d and c are directly related and differ by a known multiplicative constant $(1-p)$ only, either of these quantities serves the purpose of measuring the extent of dishonest responding equally well.[2] Because of this interchangeability (and because d is more tractable), we only consider the proportion of dishonest individuals (d) in the present article.

Estimating correlations involving dishonesty

As we sketched above, cheating paradigms implement a situation that can be conceived of as observing a response variable (alleged wins) that is a mixture of a dishonesty variable and a disturbance term that is due to actual wins by honest individuals. Stated differently, honest individuals winning the lottery (with known probability p) add a mask with known mean and zero variance (due to the assumption that each factually winning participant also reports a win) to the responses by dishonest individuals. A sample estimate of the correlation between the unobserved dishonesty variable d and an observed covariate x can be obtained from the observed "win" and "not win" responses w through

$$r_{x,d} = r_{x,w} \frac{s_w}{s_d (1-p)}, \tag{5}$$

where s_w^2 is the variance of the observed "win" responses. The estimate of the sample variance of the unobserved dishonesty variable is given by

$$s_d^2 = \frac{s_w^2 - p(1-p) \cdot \left(\hat{d}-1\right)^2}{(1-p)} \tag{6}$$

(Fox & Tracy, 1984; Himmelfarb, 2008; Kraemer, 1980). The second term in Eq. 5 serves to correct the correlation between the covariate and the observed "win" responses, $r_{x,w}$, for the attenuation due to the disturbance process (akin to the correction for attenuation due to unreliability in classical test theory). This correction factor increases with the probability of winning, so the uncorrected correlation $r_{x,w}$ underestimates the true correlation to the extent that actual winning is increasingly likely. In contrast, as p approaches 0, the variance of dishonesty becomes equal to the observed variance of the "win" responses, and the correction factor becomes equal to 1. Clearly, if one cannot actually win, any "win" response must be illegitimate, and thus stem from a dishonest individual. Since $r_{x,d}$ is 0 if and only if the correlation between the covariate and the observed "win" responses is 0 ($r_{x,w}$ = 0), it suffices to test whether $r_{x,w}$ differs from 0 (by means of a simple t test for correlations), to test whether $r_{x,d}$ differs from 0 (Kraemer, 1980).

Predicting dishonesty using logistic regression analysis

Because applications of cheating paradigms yield a dichotomous outcome variable, logistic (rather than ordinary least squares) regression is the appropriate analysis strategy. In the standard logistic regression model, the probability of a dichotomous outcome variable y with values (0, 1) is regressed on predictor variables $\mathbf{X}$ through a logistic link function,

$$p(y = 1|\mathbf{X}) = \frac{\exp(\mathbf{X}'\mathbf{B})}{1+\exp(\mathbf{X}'\mathbf{B}')}, \tag{7}$$

where $\mathbf{B}$ contains the regression parameters. In the present context, an observed alleged "win" response might also be due to actual wins by honest individuals, so that the probability of observing a "win" response is $p(\text{"win"}) = d + (1-d) \cdot p = p + (1-p) \cdot d$. The modified logistic regression model accommodating this difference becomes

$$p(\text{"win"}|\mathbf{X}) = p + (1-p) \cdot \frac{\exp(\mathbf{X}'\mathbf{B})}{1+\exp(\mathbf{X}'\mathbf{B}')} \tag{8}$$

and can be estimated by standard maximization routines (for details, see Scheers & Dayton, 1988; van den Hout, van der Heijden, & Gilchrist, 2007). The resulting regression coefficients can be interpreted as usual—that is, as the change in the logit of dishonesty given a unit change in the predictor. Similarly, exp($\mathbf{B}$) can be interpreted in the usual way as the odds ratio (OR).

[1] An equivalent test would be to use a z-test comparing the observed proportion of alleged wins with the baseline probability of winning.

[2] There might be cases in which some variable is hypothesized to be a consequence of the actual behavioral act of having cheated, but not of the principal willingness to cheat if required (dishonesty). In such an ex-post-facto situation, it would be inappropriate to rely on the estimated proportion of dishonest individuals, because only a subset of these would have cheated in the sense of reporting an outcome that they did not obtain. However, in any other situation, the willingness to cheat is directly related to the behavioral act of cheating, so that either of these measures can be used interchangeably.

Simulation studies

We performed two simulation studies to demonstrate the superiority of the proposed methods over the mere consideration of the observed rate of alleged wins.[3] The first simulation was based on a population of N = 1,000 participants comprising a constant proportion of either 10 %, 30 %, or 50 % dishonest respondents. The population data included two continuous background covariates (e.g., some personality scale or the like). The first covariate was correlated by either ρ = .25 or ρ = .50 with the true honest versus dishonest state. The second covariate was created such that its true logistic regression slope when predicting dishonesty was either β = 0.50 (OR = 1.65) or β = 1.00 (OR = 2.71). The observed "win" responses were generated by applying five different baseline probabilities of winning (ranging from p = .10 to .50). The final stage was replicated 500 times.

The mean estimated correlations between the response variable and the covariate across the 500 replications are shown in Table 1.[4] As can be seen, the true underlying correlation was consistently and substantially underestimated when simply considering the "win" responses. As expected, this bias increased with the baseline probability of winning and decreased with the proportion of dishonest individuals, because both factors increase the distortion of the observed "win" responses by honest and factually winning participants. For example, with a baseline probability of winning of p = .50 and 10 % dishonest individuals, the mean correlation estimate given a true correlation of ρ = .25 was just r = .07. As is shown in Table 2, the same effect necessarily also occurs when predicting the "win" responses (rather than dishonesty) through a logistic regression. Simply predicting the "win" responses substantially underestimates the true relationships, especially when the baseline probability of winning is high and the overall proportion of dishonest respondents is small.

In contrast to these findings, the modified procedures provided estimates that matched the true population parameters very closely. Unlike the correlation between the observed "win" responses and the background variable, the means of the corrected correlation coefficients nearly perfectly reproduced the true correlations in all conditions (Table 1), regardless of the baseline probability and the proportion of dishonest individuals. Similarly, when applying the modified logistic regression model to the simulated data, the true slopes for the covariate used to predict dishonesty were adequately recovered across conditions (Table 2). The performance of the modified logistic regression model was better when there was a larger proportion of dishonest individuals. This is to be expected, however, given that it is more difficult to predict dishonesty when only a few observations exhibiting the behavior of interest are available.

In summary, the results of the first simulation study clearly showed that (a) merely considering the observed "win" responses leads to a bias, in that the true relationships are substantially underestimated, and (b) the modified methods outlined herein are able to provide a good recovery of the true population parameters. However, the first simulation study considered only a single (large) sample size of participants in order to demonstrate that differences between the traditional and modified approaches are not due to sampling fluctuations. To investigate the performance of the methods at smaller—and more realistic—sample sizes, a second simulation study was performed in which the sample size was systematically varied between N = 25 and N = 250. The general setup was similar to that in the first study, in that a true correlation of ρ = .50 for the first covariate and a true logistic regression slope of β = 1.00 for the second covariate were used. The proportion of dishonest participants was 30 %, and the baseline probability of winning was p = .25.

Figures 1 and 2 show the results for the estimated correlations and the estimated logistic regression slopes, respectively. Replicating the first simulation study, the general pattern of results shows that the traditional analyses using the observed alleged "win" responses underestimated the true relationships, regardless of sample size. Moreover, the corrected correlation coefficients reliably recovered the true correlations across all considered sample sizes and only exhibited a very slight tendency to overestimate the true correlation when the sample was very small. Similar results were obtained concerning the modified logistic regression model. Overall, the true slope was closely recovered, although the slopes tended to become somewhat unstable when the sample was small. However, the estimates tended to stabilize when $N > 100$ and closely approached the true value. In addition, even for smaller samples, the modified procedure clearly outperformed the traditional analysis.

Empirical examples

We reanalyzed three published data sets to illustrate the relevance of the methods advocated herein for substantive work. In an online study involving N = 134 participants, Zettler, Hilbig, Moshagen, and de Vries (2015) investigated the hypothesis that dishonesty is negatively (rather than positively) related to scores on an impression management (IM) scale. To obtain a behavioral measure of dishonesty, participants were instructed to toss a coin exactly twice and were rewarded with a monetary gain if they reported exactly two successes. The baseline probability of winning thus was p = .25. Fifty-one (38 %) "win" responses were observed. The proportion of

[3] The simulation data and associated R scripts are available at https://osf.io/wg5rc/.

[4] Correlation coefficients were transformed to Fisher Z-scores, averaged, and then back-transformed.

Table 1 Mean uncorrected (r) and corrected ($\tilde{r}$) correlation estimates

	10 % Dishonest		30 % Dishonest		50 % Dishonest	
	r	$\tilde{r}$	r	$\tilde{r}$	r	$\tilde{r}$
$\rho = .25$						
$p = .100$	.17 (.02)	.25 (.04)	.21 (.02)	.25 (.02)	.23 (.01)	.25 (.02)
$p = .167$	.14 (.02)	.25 (.05)	.20 (.02)	.25 (.03)	.22 (.02)	.25 (.02)
$p = .250$	.12 (.03)	.26 (.06)	.17 (.02)	.25 (.03)	.19 (.02)	.25 (.03)
$p = .333$	.10 (.03)	.25 (.07)	.15 (.03)	.25 (.04)	.18 (.02)	.25 (.03)
$p = .500$	.07 (.03)	.26 (.10)	.12 (.03)	.25 (.06)	.14 (.02)	.25 (.04)
$\rho = .50$						
$p = .100$	.34 (.02)	.50 (.03)	.43 (.02)	.50 (.02)	.45 (.01)	.50 (.01)
$p = .167$	.29 (.02)	.50 (.04)	.39 (.02)	.50 (.02)	.42 (.02)	.50 (.02)
$p = .250$	.24 (.03)	.52 (.06)	.35 (.02)	.50 (.03)	.39 (.02)	.50 (.02)
$p = .333$	.21 (.03)	.52 (.07)	.31 (.02)	.50 (.04)	.36 (.02)	.50 (.03)
$p = .500$	.15 (.03)	.53 (.08)	.24 (.02)	.50 (.05)	.29 (.02)	.50 (.04)

Means and standard deviations in parenthesis across 500 replications. p = Baseline probability of winning.

dishonest respondents was estimated at $\hat{d} = .17$, which is significantly larger than zero, $G^2(1) = 11.1$, $p < .01$. To test the hypothesis concerning the relationship between dishonesty and IM, a logistic regression predicting the "win" responses was performed. The bivariate correlation between IM and winning was $r_{IM,w} = -.17$, and the odds ratio for IM in the logistic regression predicting the "win" responses was $OR = 0.56$ (both $p < .05$). Given the comparatively low proportion of dishonest respondents in the sample, notably different estimates would be expected when using the modified procedures described above. Indeed, the correlation between dishonesty and IM was estimated at $r_{IM,d} = -.29$, and the odds ratio for IM in the regression predicting dishonesty was $OR = 0.16$ (both $p < .05$). Thus, the original analyses relying on the observed "win" responses strongly underestimated the relationship between dishonesty and IM. Whereas the original effect would have been considered small to moderate in size, the present estimate reflects a very large effect in terms of the effect size conventions for odds ratios proposed by Rosenthal (1996).

Importantly, the modified logistic regression model described above is of course not limited to a single predictor. Hilbig and Zettler (2015, Exp. 2) used a cheating

Table 2 Mean uncorrected (b) and corrected ($\tilde{b}$) logistic regression slopes

	10 % Dishonest		30 % Dishonest		50 % Dishonest	
	b	$\tilde{b}$	b	$\tilde{b}$	b	$\tilde{b}$
$\beta = 0.50$						
$p = .100$	0.26 (0.06)	0.53 (0.12)	0.40 (0.04)	0.49 (0.04)	0.45 (0.03)	0.50 (0.03)
$p = .167$	0.20 (0.06)	0.54 (0.15)	0.35 (0.04)	0.48 (0.06)	0.42 (0.04)	0.50 (0.04)
$p = .250$	0.15 (0.06)	0.54 (0.21)	0.31 (0.05)	0.48 (0.07)	0.39 (0.05)	0.50 (0.06)
$p = .333$	0.13 (0.06)	0.56 (0.26)	0.27 (0.05)	0.47 (0.09)	0.37 (0.05)	0.50 (0.07)
$p = .500$	0.09 (0.06)	0.55 (0.41)	0.23 (0.06)	0.47 (0.14)	0.32 (0.06)	0.49 (0.10)
$\beta = 1.00$						
$p = .100$	0.50 (0.06)	1.10 (0.13)	0.76 (0.05)	0.99 (0.06)	0.86 (0.04)	0.99 (0.04)
$p = .167$	0.39 (0.06)	1.13 (0.17)	0.66 (0.05)	1.00 (0.07)	0.80 (0.05)	0.99 (0.06)
$p = .250$	0.30 (0.06)	1.17 (0.22)	0.57 (0.05)	1.01 (0.10)	0.72 (0.05)	0.99 (0.07)
$p = .333$	0.24 (0.06)	1.18 (0.26)	0.50 (0.05)	1.00 (0.11)	0.66 (0.05)	1.00 (0.09)
$p = .500$	0.17 (0.06)	1.17 (0.41)	0.39 (0.05)	1.00 (0.15)	0.57 (0.06)	0.99 (0.12)

Means and standard deviations in parenthesis across 500 replications. p = Baseline probability of winning.

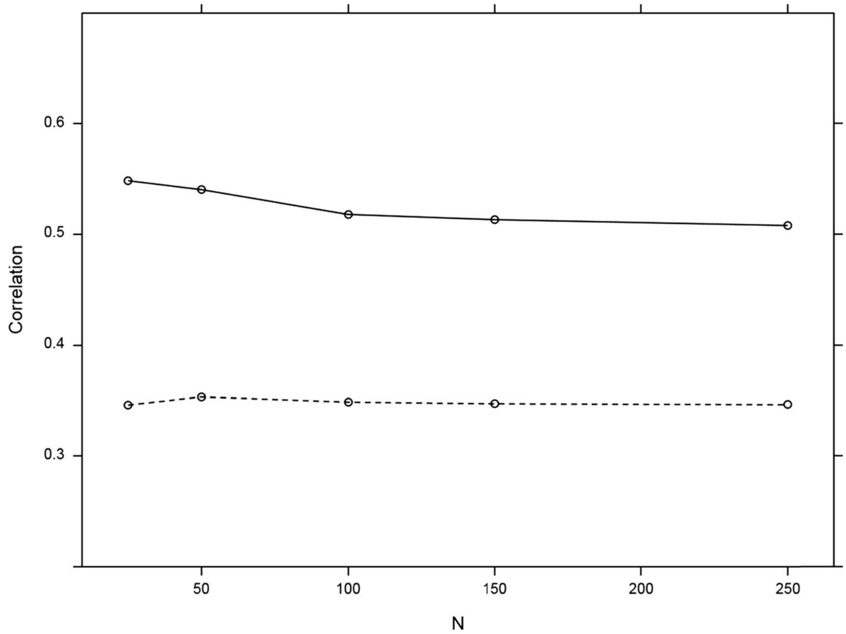

Fig. 1 Mean estimated uncorrected (*dashed line*) and corrected (*solid line*) correlations between dishonesty and a background variable as a function of sample size. The true correlation was $\rho = .50$

paradigm to investigate the relationship between dishonesty and six personality factors from the HEXACO model of personality (Ashton & Lee, 2007; Ashton, Lee, & De Vries, 2014; Moshagen, Hilbig, & Zettler, 2014). They specifically predicted that the honesty–humility personality factor would account for cheating over and above the remaining five personality factors of the model. Participants rolled a fair die in secret and obtained a monetary reward if they reported having obtained a target number, so that the baseline probability of winning was $p = 1/6$. Out of the $N = 88$ participants, 40 (45 %) allegedly won. The estimated proportion of dishonest individuals was $\hat{d} = .35$, $G^2(1) = 39.5$, $p < .01$. To evaluate their hypotheses, Hilbig and Zettler (2015) performed a logistic

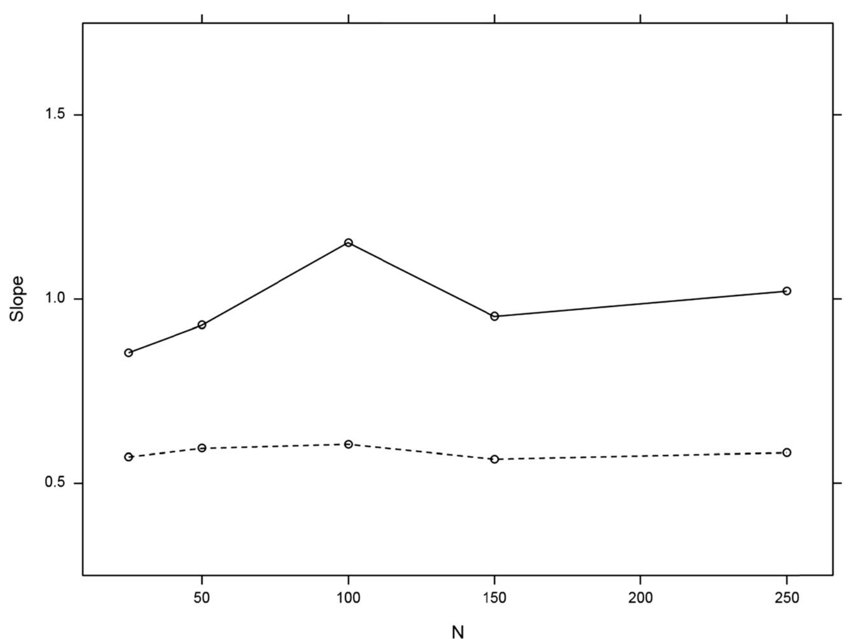

Fig. 2 Mean estimated slopes for the prediction of dishonesty by a background variable obtained from traditional (*dashed line*) and modified (*solid line*) logistic regression analyses, as a function of sample size. The true regression slope was $\beta = 1.00$

regression analysis predicting the observed "win" responses from the six HEXACO factors, which yielded an odds ratio for the honesty–humility factor of OR = 0.32 (p < .05). Using the modified logistic regression model, the odds ratio for the honesty–humility factor in the prediction of dishonesty by the six HEXACO factors was OR = 0.28 (p < .05), which is similar to the OR obtained when predicting the "win" responses. In this study, the results of the traditional and modified methods are highly similar due to a rather low baseline probability of winning, in combination with a substantial proportion of dishonest respondents, in turn implying only a negligible contamination of the "win" responses by individuals legitimately winning.

To illustrate the benefits of the suggested methods for experimental studies, we finally consider an experiment by Hilbig and Hessler (2013), who used a dice-rolling paradigm to test whether the willingness to cheat depends on how great a lie would be required to obtain the gain (as is predicted by theories focusing on self-concept maintenance; cf. Mazar et al., 2008). A total of N = 765 participants rolled a fair, six-sided die in secret. The target number (the one outcome incurring a gain) was varied experimentally, but the winning probability was always $p = 1/6$. Across conditions, the proportion of dishonest respondents was $\hat{d} = .23$, $G^2(1) = 164.9$, $p < .01$. Hilbig and Hessler formulated the hypothesis that dishonesty would be least prevalent for the target numbers 1 and 6 (since these provide the least opportunity for relatively minor lies), intermediate for 2 and 5, and most prevalent for 3 and 4 (since these provide the most opportunities for more minor lies). Using target numbers of 1 and 6 as the reference category, a standard logistic regression predicting the "win" responses by dummy-coded predictor variables yielded OR = 1.92 for the targets 2 and 5 and OR = 2.40 for the targets 3 and 4 (both ps < .01), thus indicating small to moderate effects. In contrast, applying the modified logistic regression model to predict dishonesty yielded large effects, with estimated odds ratios of OR = 3.35 for the targets 2 and 5 and OR = 4.54 for the targets 3 and 4 (both ps < .01).

Discussion and conclusion

The field of behavioral ethics has seen an upsurge of interest in recent years, and paradigms suited to study dishonesty have correspondingly gained much prominence (Bazerman & Gino, 2012; Rosenbaum et al., 2014). One type of paradigm essentially implements a lottery, relying on a randomization device with a known probability distribution (such as coins, dice, etc.) to determine winnings. However, the true outcome of the randomization process is only known to the participants, who can thus cheat by claiming to have won, despite having obtained a different outcome. Data obtained in such a way are typically analyzed using the observed "win" or "not win" responses. As a consequence, hypotheses involving dishonesty are only tested indirectly, and more importantly, correlations between covariates and dishonesty, as well as regression coefficients in the prediction of dishonesty, are—sometimes seriously—underestimated (as we have shown in the simulation studies above). As a remedy, we have outlined approaches to estimate correlations and predict dishonesty, rather than using the observed proportion of alleged wins as a proxy. In simulations we showed that the adapted methods allow for obtaining undistorted estimates of the correlations and logistic regression coefficients. Also, by reanalyzing empirical data, we demonstrated the relevance of the described methods for substantive applications.

Despite these advantages over simply using the observed proportion of alleged wins, the modified procedures also bear a caveat: When predicting dishonesty, honest participants who actually won essentially constitute random noise contaminating the response variable. This is desirable, because it renders the paradigm truly anonymous (i.e., it is not self-incriminating to claim to have won). However, the immediate consequence of this contamination is that standard errors are larger than those obtained with uncontaminated variables. One obvious way to counter this effect is to increase the sample size. Other than that, certain features of the experimental setup can be used to narrow the confidence limits. In particular, choosing high baseline probabilities of winning (say, p = .5) will amplify this problem, so rather low baseline probabilities are to be preferred (with respect to smaller standard errors). At the same time, however, cheating must be an attractive option, so as to avoid only having a few participants cheat. As p approaches 0, "win" responses become more incriminating, thus making cheating less likely, so p should be sufficiently larger than 0 to render a "win" response unsuspicious. On the basis of the (limited) simulations presented herein, a winning probability of p = .25 may serve as a reasonable compromise. Note, however, that the undistorted effect size estimates obtained by the adapted methods (which must be larger—as long a true effect exists—than those obtained using the "win" response) offset the larger standard errors.

Paradigms suited to study dishonest behavior are of quintessential importance in the field of behavioral ethics to examine the determinants, boundary conditions, and underlying processes of unethical behavior. To this end, the methods outlined herein allow for directly considering dishonesty in different analytical frameworks. Thus, by overcoming the limitations inherent in merely considering reported wins, these methods will provide a considerable benefit for future research involving cheating paradigms.

Appendix

The approaches to analyze data obtained from cheating paradigms outlined in the present article can be performed in the R environment using the "RRreg" package (available from CRAN). In the following code examples, the observed "not win"/"win" responses (coded by 0 and 1, respectively) are stored in a vector called r. The baseline probability of winning is denoted p. For illustration, we refer to a study that used a baseline probability of $p = .25$ and observed 595 participants who reported not having won ($r_i = 0$) and 405 participants claiming to have won ($r_i = 1$), so that the observed proportion of alleged wins was $q = 40.5$ %. The associated R script can be downloaded from https://osf.io/wg5rc.

The proportion of dishonest respondents can be estimated via the function RRuni.

```
p <- .25
res.uni <- RRuni(r, model = 'FR', p = c(0, p))
summary(res.uni)
```

The (shortened) output is:

```
>      Estimate  StdErr     z   Pr(>|z|)
> pi0  0.793    0.021   38.329  < 2.2e-16 ***
> pi1  0.207    0.021    9.985  < 2.2e-16 ***
```

The first row (pi0) gives the estimated proportion (along with associated standard error) of honest participants, the second row (pi1) the proportion of dishonest participants.

To estimate the correlation between dishonesty and another (noncontaminated) variable x, the function RRcor is used.

```
p <- .25
RRcor(r, x, models = c('FR', 'd'), p.list = list(c(0,p)))
```

The function RRlog computes a logistic regression with a contaminated dependent variable. The general syntax is similar to R's glm command. Assume that dishonesty is predicted by two uncontaminated (continuous or dichotomous) variables x1 and x2,

```
p <- .25
res.log <- RRlog(r ~ x1 + x2, model = 'FR', p = c(0,p))
summary(res.log)
```

The (shortened) output is:

```
>              Estimate  StdErr  deltaG2  Pr(>deltaG2)
> (Intercept)  -2.46     0.29    293.85        < 2e-16
> x1            1.50     0.19    121.92        < 2e-16
> x2           -0.61     0.15     19.63          1e-05
```

So, the logistic regression coefficients when predicting dishonesty for the predictor variables x1 and x2 are 1.50 and –0.61, respectively. All regression coefficients differ significantly from 0. To obtain odds ratios, the command `exp (coef(res.log))` can be used. This shows that the `OR` for a unit change in x1 is `OR` = 4.50, and for x2 is `OR` = 0.54.

References

Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics, 113,* 96–104.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150–166.

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty–Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18,* 139–152.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science, 8,* 85–104.

Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General, 142,* 1001–1005.

Caruso, E. M., & Gino, F. (2011). Blind ethics: Closing one's eyes polarizes moral judgments and discourages dishonest behavior. *Cognition, 118,* 280–285. doi:10.1016/j.cognition.2010.11.008

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34,* 1–7.

Fischbacher, U., & Heusi, F. (2008). *Lies in disguise, an experimental study on cheating*. Research Paper Series, Thurgau Institute of Economics, University of Konstanz. Retrieved from: http://www.twikreuzlingen.ch/uploads/tx_cal/media/TWI-RPS-040-Fischbacher-Heusi-2008-11.pdf.

Fischbacher, U., & Utikal, V. (2011). *Disadvantageous lies*. Research Paper Series, Thurgau Institute of Economics, University of Konstanz. Retrieved from: http://www.twi-kreuzlingen.ch/uploads/tx_cal/media/TWIRPS-071-Fischbacher-Utikal-2011-11.pdf.

Fox, J. A., & Tracy, P. E. (1984). Measuring associations with randomized response. *Social Science Research, 13,* 188–197.

Gino, F., & Margolis, J. D. (2011). Bringing ethics into focus: How regulatory focus and risk preferences influence (un) ethical behavior. *Organizational Behavior and Human Decision Processes, 115,* 145–156.

Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes, 117,* 298–310.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology, 49,* 263–266.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behaviour. *Journal of Research in Personality, 57,* 72–88.

Himmelfarb, S. (2008). The multi-item randomized response technique. *Sociological Methods & Research, 36,* 495.

Kraemer, H. C. (1980). Estimation and testing of bivariate association using data generated by the randomized response technique. *Psychological Bulletin, 87,* 304–308.

Lewis, A., Bardis, A., Flint, C., Mason, C., Smith, N., Tickle, C., & Zinser, J. (2012). Drawing the line somewhere: An experimental study of moral compromise. *Journal of Economic Psychology, 33,* 718–725.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45,* 633–644.

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing, 25,* 117–126.

Mazur, M. J., & Plumley, A. H. (2007). Understanding the tax gap. *National Tax Journal, 60,* 569–576.

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology, 45,* 594–597.

Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). A strategy for validating questioning techniques that assess sensitive issues. *Experimental Psychology, 61,* 48–54.

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO Personality Inventory]. *Diagnostica, 60,* 86–97.

Nogami, T., & Yoshida, F. (2013). Rule-breaking in an anonymous situation: When people decide to deviate from existing rules. *International Journal of Psychology, 48,* 1284–1290.

Peer, E., Acquisti, A., & Shalvi, S. (2014). "I cheated, but only a little": Partial confessions to unethical behavior. *Journal of Personality and Social Psychology, 106,* 202–217.

Read, T., & Cressie, N. A. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York, NY: Springer.

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

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21,* 37–59.

Scheers, N. J., & Dayton, C. M. (1988). Covariate randomized response models. *Journal of the American Statistical Association, 83,* 969–974.

Schurr, A., Ritov, I., Kareev, Y., & Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgment and Decision Making, 7,* 679–688.

Schweitzer, M. E., Ordóñez, L., & Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Academy of Management Journal, 47,* 422–432.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23,* 1264–1270. doi:10.1177/0956797612443835

Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management, 22,* S16–S27.

Van den Hout, A., van der Heijden, P. G., & Gilchrist, R. (2007). The logistic regression model with response variables subject to randomized response. *Computational Statistics and Data Analysis, 51,* 6060–6069.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will encouraging a belief in determinism increases cheating. *Psychological Science, 19,* 49–54.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences, 112,* 10651–10656.

Zettler, I., Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue? A behavioral test of impression management. *Personality and Individual Differences, 81,* 107–111.

Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science, 21,* 311–314. doi:10.1177/0956797609360754

Zimerman, L., Shalvi, S., & Bereby-Meyer, Y. (2014). Self-reported ethical risk taking tendencies predict actual dishonesty. *Judgment and Decision Making, 9,* 58–64.