Received: 10 October 2019 | Revised: 5 June 2020 | Accepted: 5 June 2020
DOI: 10.1002/bdm.2195

RESEARCH ARTICLE

WILEY

Behavior in cheating paradigms is linked to overall approval rates of crowdworkers

Christoph Schild[1,2] | Lau Lilleholt[1] | Ingo Zettler[1]

[1]Department of Psychology, University of Copenhagen, Copenhagen, Denmark

[2]Department of Psychology, University of Siegen, Siegen, Germany

Correspondence
Christoph Schild, Department of Psychology, University of Siegen, Adolf-Reichwein-Str. 2a, Siegen 57068, Germany.
Email: christoph-schild@uni-siegen.de

Funding information
Carlsbergfondet, Grant/Award Number: CF16-0444; Det Frie Forskningsråd, Grant/Award Number: 7024-00057B

Abstract

Dishonest and fraudulent behavior poses a serious threat to both individuals and societies. Many studies investigating dishonesty rely on (one of) a few well-established lab and online cheating paradigms. Quite surprisingly, though, the external validity of these paradigms has only been investigated in a small number of studies, raising the question of whether behavior in these paradigms is related to real-life dishonesty or, more broadly, socially questionable behavior. Tackling this gap, we link observed behavior in two widely used cheating paradigms to approval rates on two crowdworking platforms (namely, Prolific and Amazon Mechanical Turk) using data from four studies (overall N = 5,183). Results indicate that lower approval rates are associated with higher proportions of dishonest individuals. Importantly, this relation also holds for crowdworkers who exceed commonly used thresholds for study inclusion. The results thus support the external validity of (two widely used) cheating paradigms. Further, the study identifies approval rates as a variable that explains dishonesty on crowdworking platforms.

KEYWORDS

cheating, coin flip, crowdworking, dishonesty, external validity, Mind Game

1 | INTRODUCTION AND THEORETICAL BACKGROUND

Individuals and societies are constantly affected by dishonest and fraudulent behavior. Dishonesty comes in many forms, including recent large-scale examples of emission cheating scandals (Volkswagen, 2015), money laundering (Danske Bank, 2018), and systematic college admission frauds (Thelin, 2019). Irrespective of the kind of dishonest behavior, most acts of dishonesty undermine interpersonal and/or societal well-functioning and can have tremendous negative consequences (Del Monte & Papagni, 2001; Gyimah-Brempong, 2002; Judge, McNatt, & Xu, 2011; Mo, 2001).

In the last years, many studies investigating the occurrence and extent of dishonesty as well as its predictors, correlates, and consequences have used (variants of) a few well-established cheating paradigms, which are conceptually quite similar to each other. In a recent meta-analysis on dishonesty, for instance, Gerlach, Teodorescu, and Hertwig (2019) considered (variants of) four different cheating paradigms, namely, the coin flip paradigm (Bucciol & Piovesan, 2011), the die roll paradigm (Fischbacher & Föllmi-Heusi, 2013), the matrix task (Mazar, Amir, & Ariely, 2008), and the sender-receiver game (Gneezy, 2005). In each of these paradigms, participants have the opportunity to act dishonestly in order to obtain an incentive.

In the coin flip paradigm, for instance, participants are asked to flip a coin in private and to report their outcome (i.e., "heads" or "tails"). Typically, the report of a specific outcome is incentivized (e.g., a participant earns $1 for reporting "heads"), making it possible for a participant to misreport their outcome in order to obtain the

TABLE 1 Overview of studies linking behavior in cheating paradigms to real-world socially questionable behavior

Study	N	Cheating paradigm subgroup(see Gerlach et al., 2019)	Real-world outcome	Finding
Cohn, A., & Maréchal, M. A. (2018).	162 students	Coin flip paradigm	School misconduct	Cheating was positively linked to disruptiveness and homework noncompletion but not to absenteeism.
Cohn, A., Maréchal, M. A., & Noll, T. (2015).	182 prison inmates	Coin flip paradigm	Rule violation	Cheating was positively linked to the number of rule violation offenses in prison.
Dai, Z., Galeotti, F., & Villeval, M. C. (2017).	279 train riders	Die roll	Fare dodging	Non-ticket holders (fraudsters) cheat more than ticket holders (non-fraudsters).
Hanna, R., & Wang, S.-Y. (2017).	165 nurses	Die roll	Work absence	Cheating was negatively linked to attendance.
Kröll, M., & Rustagi, D. (2016).	72 milk men	Die roll	Cheating behavior in milk markets	Cheating was positively linked to more added water to milk (i.e., fraud).
Potters, J., & Stoop, J. (2016).	102 students	Mind Game	Nonreporting of overpayment	Subjects with higher payoffs in the Mind Game were less likely to report overpayment.

specified incentive. Other cheating paradigms follow a similar logic—that is, participants are given a chance to misreport the outcome of an event in a highly anonymous setting in order to obtain an incentive (or to avoid losing an advantage). Importantly, in such paradigms, it is typically not recorded (and, thus, known) whether any specific individual has cheated or not.[1] Rather, researchers draw conclusions about the proportion of dishonest individuals (and characteristics of these) by comparing the number of alleged wins for the whole sample with the stochastic baseline of winning (e.g., in the coin flip paradigm with one trial, the stochastic baseline of winning is 50%; for more details about this, see Moshagen & Hilbig, 2017).

Clearly, investigating dishonest behavior under such controlled and anonymized conditions has many advantages (e.g., protecting participants' anonymity which should also reduce socially desirable reporting of outcomes). At the same time, one might question whether people's behavior in such paradigms is externally valid and transferable to (larger-scale) real-life behavior (i.e., actual behavior outside a lab or online research studies). Surprisingly, though, comparing people's behavior in such cheating paradigms with socially questionable real-world behavior[2] has hardly been investigated to date. Indeed, to the best of our knowledge, only six studies have so far investigated how behavior in cheating paradigms relates to socially questionable real-world behavior (see Table 1). Therein, behavior in cheating paradigms has been linked to classroom misbehavior in schools (Cohn & Maréchal, 2018), offenses against prison regulation among inmates (Cohn, Maréchal, & Noll, 2015), fare dodging in public transport (Dai, Galeotti, & Villeval, 2017), absence from work among nurses (Hanna & Wang, 2017), fraudulent salesmen behavior (Kröll & Rustagi, 2016), and nonreporting of overpayment (Potters & Stoop, 2016). Overall, these findings are clearly in line with personality trait theory (Allport, 1961), which assumes that individuals do have rather stable personality characteristics that influence behaviors, thoughts, and emotions across different contexts; that is, next to situational characteristics affecting the occurrence and/or extent of certain behavior, personality trait theory would predict that some individuals are generally more likely than others to engage in certain behaviors (such as socially questionable behavior) and that the increased likelihood of engaging in a certain kind behavior can be observed across different contexts. With regard to socially questionable behavior, this assumption is well supported by meta-analytic evidence. For instance, Zettler, Thielmann, Hilbig, and Moshagen (2020) recently found that people with rather low levels in the personality dimension of Honesty–Humility tend to show not only more cheating/dishonesty but also, among other things, more aggression, antisocial behavior, counterproductive behavior, or criminality/delinquency.

Although all of the studies mentioned in Table 1 do support that behavior in cheating paradigms is a valid indicator of socially questionable real-world behavior, they are limited by the used sample sizes in particular. Specifically, with an average sample size of 161 and considering that cheating paradigms come with certain limitations of statistical power (i.e., cheating is typically unknown on the individual level),

[1]Note that in some studies, each participant is (un)knowingly observed by the experimenter, making it possible to identify honest and dishonest respondents on the individual level (e.g., Kocher, Schudy, & Spantig, 2017; Kröll & Rustagi, 2016)

[2]Because not all of the following criteria might be clearly classified as dishonesty, we use the broader term socially questionable behavior. Please note, though, that each of the following criteria (as well as the criteria in our studies) relates to dishonest behavior to some degree.

more well-powered studies are needed to test whether behavior in cheating paradigms can indeed be linked to real-world socially questionable behavior. We tackle this gap by a series of four studies.

2 | THE PRESENT INVESTIGATION

Adding to the existing literature on the external validity of cheating paradigms, we link two cheating paradigms—namely, the coin flip (Bucciol & Piovesan, 2011) and the Mind Game (Jiang, 2013) paradigm—to crowdworkers' approval rates on the crowdwork platforms Prolific (Study 1–3) and Amazon Mechanical Turk (MTurk; Study 4), respectively. Crowdworkers are workers that participate in crowdsourced tasks in exchange for an (monetary) incentive. Tasks on Prolific mostly consist of study participation. However, studies conducted on this platform are very diverse and from multiple disciplines such as clinical psychology (Alexander, Salum, Swanson, & Milham, 2020), cognitive psychology (e.g., Ensor, Surprenant, & Neath, 2019), health psychology (e.g., Todd, Aspell, Barron, & Swami, 2019), social psychology (e.g., Jolley, Douglas, Leite, & Schrader, 2019), and economics (e.g., Teubner, Hawlitschek, & Adam, 2019). Tasks on MTurk do also include study participation but do largely consist of business requests such as classification tasks, product reviews, or transcriptions (e.g., Ipeirotis, 2010).

Important for our research question is that crowdworkers' submissions are evaluated by the conducting party and have to be accepted or rejected. The number of accepted and rejected submissions is used by Prolific and MTurk, respectively, to calculate an approval rates score for each individual crowdworker; that is, if a worker gets a study accepted, the rate goes up, and if a worker gets a study rejected, the rate goes down.

Reasons for rejection are manifold but can include misrepresentation of study inclusion criteria, deception of the requester, or provision of quick random responses (e.g., Hydock, 2018; Johnstone, Tate, & Fielt, 2018; Prolific Team, 2018). It can thus be assumed that approval rates partly indicate workers' dishonesty in their past submissions on crowdworking platforms. Importantly, acting dishonestly on crowdsourcing platforms with regard to the task requirements[3] comes with an important trade-off: Crowdworkers can act dishonestly in order to save time and/or increase their financial benefit (e.g., by being able to participate in more studies in a specific timeframe), but they do also risk rejections which lower their approval rates and, in turn, might prevent them from participating in future tasks on the platform (for some tasks, requestors—i.e., the ones who conduct the tasks—set a minimum approval rate as a requirement for task participation; e.g., Ensor et al., 2019; Grysman, 2015).

Overall, crowdworkers' approval rates thus represent an indicator of real-world socially questionable behavior (with lower approval rates indicating more socially questionable behavior across numerous tasks). In line with the predictions of personality trait theory, we hypothesize that individuals with lower approval rates are more likely to cheat in a cheating paradigm than individuals with higher approval rates. While most researchers set a minimum approval rate as a requirement for study participation (e.g., a Prolific score of at least 90; Ensor et al., 2019; or an approval rate of at least 95 on MTurk; Grysman, 2015), we also investigate whether the relation holds more generally beyond commonly used thresholds for study inclusion by using a broader range of approval rates.

Next to the aim of testing the external validity of cheating paradigms, this study also allows us to potentially identify a new control variable for studies conducted on crowdworking platforms, which are often used for studying dishonesty (e.g., Gerlach et al., 2019; Peer, Brandimarte, Samat, & Acquisti, 2017; Pfattheicher & Keller, 2018). Specifically, indicators of overall submission quality, such as approval raties, might add to other characteristics that are known to influence dishonest behavior, such as age or gender (Gerlach et al., 2019), increasing the interpretability of research findings.

Overall, we present four studies investigating how approval rates are related to cheating behavior in three different cheating paradigms. In detail, Study 1 and Study 4 used an adapted version of the Mind Game (Schild, Heck, Ścigała, & Zettler, 2019), Study 2 used a standard version of the coin-flip paradigm (e.g., Hilbig & Zettler, 2015), and Study 3 used a computerized coin flip paradigm (e.g., Balasubramanian, Bennett, & Pierce, 2017). Further, across the studies we test for a broad range of approval rates (even beyond commonly used thresholds) as well as for the robustness of our findings by controlling for age and gender of the participants, as these were suggested as predictors of dishonesty in cheating paradigms in the recent meta-analysis by Gerlach et al. (2019). Studies 1–3 were conducted on Prolific, whereas Study 4 was conducted on MTurk.

3 | STUDY 1

3.1 | Methods

3.1.1 | Procedure and variables

Studies 1–3 were originally conducted for a different main purpose (and preregistered with regard to the different purpose, including a priori power calculations concerning the targeted sample size). Importantly, though, the herein reported data of Studies 1–3 represent conditions in which participants were confronted with demographic questions and a particular cheating paradigm only (i.e., no other measures or interventions were administered).

Concerning Study 1, we conducted an online experiment using the open-source survey framework formr (www.formr.org; Arslan, Walther, & Tata, 2019; Arslan & Tata, 2019). A total of 293 participants completed the experiment via Prolific. Participants were heterogeneous with respect to gender (53.92% female, 45.73% male, 0.34% other) and (although slightly less so) age (M = 36.22, SD = 12.16 years). Only participants with a Prolific score of 95 or higher were invited to the study. Prolific scores represent the upper bound of the 99th

[3]Cheating in a cheating paradigm does not lead to rejection of a crowdworker, because this behavior is in line with the study requirements (e.g., reporting an outcome).

percentile binomial confidence interval (with respect to their percentage of approved submissions from the total) and range from 0 to 100. Note that approximately 97% of the active users on Prolific (i.e., users that were active during the last 90 days) do have a Prolific score of 95 or higher (as of June 4, 2020).

Participants were informed that the main aim of the study was to investigate decision-making processes. After consenting to participate in the study, participants provided demographic information. Next, participants participated in an adapted version of the Mind Game paradigm (Schild et al., 2019). Specifically, participants were asked to write down a target number between 1 and 8 in private. Subsequently, a random number between 1 and 8 was displayed, and participants were asked whether the displayed number matched the target number they wrote down beforehand. Importantly, in addition to their flat-fee for participation (£0.40), participants received a bonus incentive of £0.40 when reporting a match. Consequently, participants had the opportunity to cheat in order to obtain the bonus incentive by reporting that the numbers matched even if they did not. Directly after the data collection was finished, approval rates (M = 99.59, SD = 0.83) were downloaded via the "export" function on Prolific.

3.1.2 | Analyses

In the cheating paradigm, the proportion of dishonest individuals d was estimated as described in Moshagen and Hilbig (2017). In contrast to analyses of binary cheating paradigms that simply compare the expected percentage of winners (which equals 12.5% in our case due to using eight random digits) with the observed proportion of winners, the modeling approach by Moshagen and Hilbig takes into account that the observed proportion of alleged wins is contaminated by honest respondents who actually won. To estimate the relation between the proportion of dishonest individuals and the approval rate scores, a modified logistic regression model was used. The described analyses were conducted using the RRreg package (Heck & Moshagen, 2018). Although our hypothesis is directional (i.e., lower Prolific scores are linked to a higher proportion of dishonest individuals d), two-tailed tests were used, because we originally conducted the study for a different purpose.

3.1.3 | Results

A total of 32.42% of the participants indicated a matching number, which is significantly different from the stochastic baseline of 12.5%, Z = 7.27, p < .001. The proportion of dishonest individuals was estimated to d = .23, SE = .03. The modified logistic regression showed that individuals with lower Prolific scores were more likely to be dishonest (estimate = –0.33, SE = 0.15, Wald test = 4.84, p = .038, OR = 0.72) as illustrated in Figure 1. Further, another modified logistic regression including age, gender, and Prolific score as predictors indicated that age (estimate = –0.54, SE = 0.26, Wald test = 4.36, p = .013, OR = 0.58) and Prolific score (estimate = –0.37, SE = 0.16, Wald test = 5.66, p = .021, OR = 0.69), but not gender (estimate = 0.24, SE = 0.37, Wald test = 0.41, p = .521, OR = 1.27), were significant predictors. This indicates that younger individuals with lower Prolific scores were more likely to be dishonest.

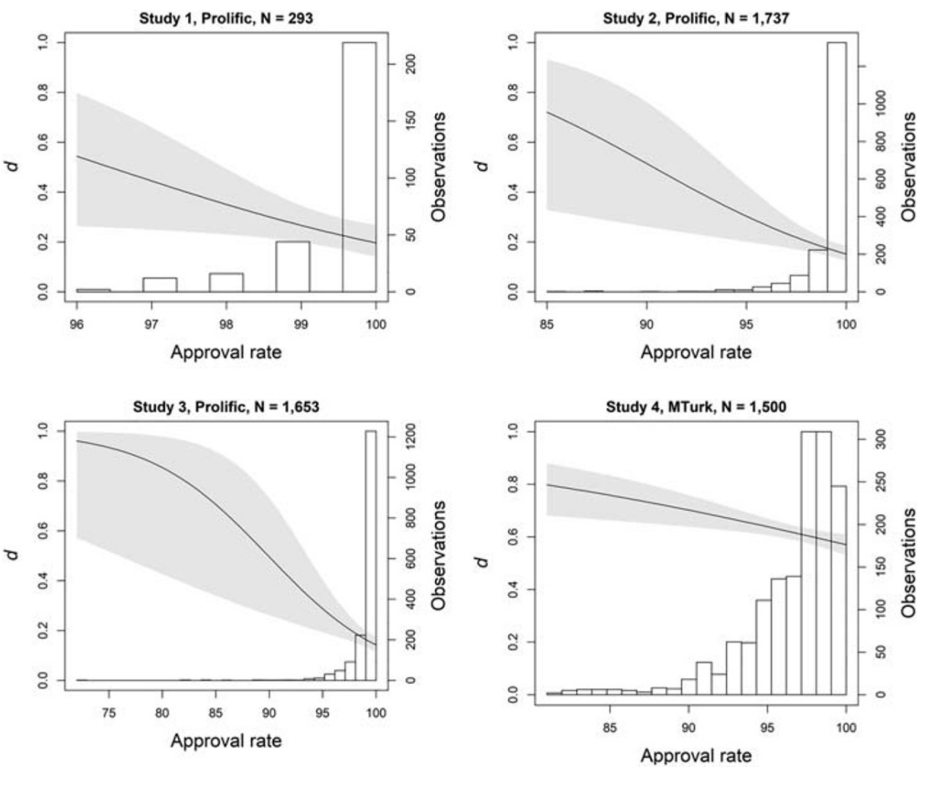

FIGURE 1 Relations between approval rates and the proportion of dishonest individuals (d) in Studies 1–4

3.1.4 | Discussion

Study 1 provided first evidence that Prolific scores are linked to cheating behavior, as assessed via the Mind Game. Study 2 tested whether these findings apply to a different cheating paradigm, namely, the coin flip task. Further, Study 2 uses a broader range of Prolific scores and a larger sample.

4 | STUDY 2

4.1 | Methods

4.1.1 | Procedure and variables

We again conducted an online experiment using the open-source survey framework formr (www.formr.org; Arslan & Tata, 2019; Arslan et al., 2019) that was originally set up with regard to a different research question (for details, see Lilleholt, Schild, & Zettler, 2020). Specifically, overall, N = 1,737 participants completed the same cheating task, though at one of two measurement occasions (2 weeks apart); that is, 867 participants completed the cheating task at the first measurement occasion, and 914 different participants completed the same cheating task at the second measurement occasion. There was no difference in the experimental setup between the two measurement occasions (i.e., we run the exact same study, just with 2 weeks apart), so that we merged these participants. Two measurement occasions were also not initially planned but had to be done because of some technical problems during the first measurement occasion. However, these did not influence the data (i.e., the conditions) reported herein. Forty-four participants had previously participated in Study 1 and were thus not included in the analyses. However, including them did not change the pattern of the results.

Participants were relatively heterogeneous with respect to gender (61.49% female, 37.82% male, 0.69% other) and age (M = 36.02, SD = 12.36 years). In the experiment, participants were first informed about the background of the study, following by providing consent and demographic information. Next, the participants were asked to play a standard version of the coin flip task as used by Zettler, Hilbig, Moshagen, and de Vries (2015). In this version of the coin flip task, participants were asked to flip a real coin twice and report the outcome in private. If participants reported flipping two heads in a row, they received a monetary payoff of £0.40, in addition to their flat fee for participation (£0.40).

For this data, we downloaded approval rates (M = 99.51, SD = 1.27) via the "export" function on Prolific approximately 4 months after the experiment (second measurement occasion) had been conducted.[4] In contrast to Study 1, data also include participants whose approval rates were lower than 95 (namely, between 85 and 100), at the time when this information was downloaded (when the experiment was launched, the required approval rate for participation was 95).

4.1.2 | Results

The same analytical approach as in Study 1 was used. However, note that the expected percentage of winners was 25% in this study (mirroring the probability of flipping two heads in a row). A total of 37.42% of the participants reported flipping two heads in a row, which is significantly different from the stochastic baseline of 25%, Z = 10.69, p < .001. The proportion of dishonest individuals was estimated to d = .17, SE = .02. The modified logistic regression showed that individuals with lower Prolific scores were more likely to be dishonest (estimate = −0.23, SE = 0.08, Wald test = 8.81, p = .010, OR = 0.80) as illustrated in Figure 1. Further, another modified logistic regression including age, gender, and Prolific score as predictors indicated that age (estimate = −0.49, SE = 0.15, Wald test = 11.57, p < .001, OR = 0.61), gender (estimate = 0.65, SE = 0.23, Wald test = 8.18, p = .004, OR = 1.92), but not Prolific score (estimate = −0.18, SE = 0.10, Wald test = 3.58, p = .092, OR = 0.83), were significant predictors. This indicates that younger, male individuals were more likely to be dishonest.

4.1.3 | Discussion

Based on a much larger sample, Study 2 conceptually replicated the findings of Study 1 using a different cheating paradigm. Unlike in Study 1, however, when age and gender were controlled for, Prolific scores did not turn out to be a significant predictor of the proportion of dishonest individuals. We ran a third study again alternating the administered cheating paradigm (this time, a computerized coin flip paradigm was used) in order to further investigate the generalizability of whether Prolific scores are linked to cheating behavior.

5 | STUDY 3

5.1 | Methods

5.1.1 | Procedure and variables

We again conducted an online experiment using the open-source survey framework formr (www.formr.org; Arslan & Tata, 2019; Arslan et al., 2019) that was originally set up with regard to a different research question (for details, see Lilleholt et al., 2020). Overall, 1,653 participants completed the experiment. Similar to Study 2, the same cheating task was administered at two measurement occasions, approximately 2 weeks apart; that is, 872 participants took part at the

[4]For the first measurement occasion in Studies 2 and 3, Prolific scores were also downloaded directly after the study was conducted. However, Prolific scores were significantly linked to dishonesty no matter which Prolific scores were used. For simplicity, we thus report results with Prolific scores that were downloaded after 4 months. We provide additional analyses in the supplementary material (https://osf.io/w8csa/?view_only=d800580e51704dcd870db8ac2e0a5540).

first measurement occasion, whereas 823 different participants took part at the second measurement occasion. There was no difference in the experimental setup between the two measurement occasions. Forty-two participants had previously participated in Study 1 and were thus not included in the analyses. However, including them did not change the pattern of the results.

Participants were relatively heterogeneous with respect to gender (59.23% female, 39.87% male, 0.91% other) and age (M = 35.85, SD = 12.07 years). After consenting to participate in the study, participants provided demographic information. Next, the participants were asked to play a computerized version of the coin flip task as in Study 2. In this version of the coin flip task, the participants were asked to flip a virtual coin via an external provider (https://justflipacoin.com) twice and report the outcome. If a participant reported that the virtual coin showed two heads in a row, they received a monetary payoff of £0.40, in addition to their flat fee for participation (£0.40). Approval rates (M = 99.43, SD = 1.61) were downloaded via the "export" function on Prolific approximately 4 months after the experiment (second measurement occasion) had been conducted. Again, we did also include participants whose approval rates were lower than 95 (range 72–100) at the time when this information was downloaded (when the experiment was set up, only crowdworkers with an approval rate of min. 95 were allowed to participate).

5.1.2 | Results

The same analytical approach as in Studies 1 and 2 was implemented. As in Study 2, the expected percentage of winners was 25%. A total of 36.60% of the participants indicated observing two heads in a row, which is significantly different from the stochastic baseline of 25%, Z = 9.79, p < .001. The proportion of dishonest individuals was estimated to d = .15, SE = .02. The modified logistic regression showed that individuals with lower Prolific scores were more likely to be dishonest (estimate = −0.29, SE = 0.09, Wald test = 10.64, p = .001, OR = 0.75) as illustrated in Figure 1. Further, another modified logistic regression including age, gender, and Prolific score as predictors indicated that gender (estimate = 0.60, SE = 0.24, Wald test = 5.07, p = .024, OR = 1.81) and Prolific score (estimate = −0.24, SE = 0.09, Wald test = 6.81, p = .007, OR = 0.78), but not age (estimate = −0.22, SE = 0.15, Wald test = 2.05, p = .120, OR = 0.80), were significant predictors. This indicates that male individuals with lower Prolific scores were more likely to be dishonest.

5.1.3 | Discussion

Study 3 conceptually replicated the findings of Studies 1 and 2 using a different implementation of a cheating paradigm (namely, via an external panel provider). We ran a final study on a different crowdworking platform—MTurk—in order to further test the generalizability of the results.

6 | STUDY 4

6.1 | Methods

6.1.1 | Procedure and variables

We again conducted an online experiment using the open-source survey framework formr (Arslan et al., 2019; Arslan & Tata, 2019). In contrast to Studies 1–3, Study 4 was specifically conducted to investigate the relation between approval rates and dishonest behavior and correspondingly preregistered (https://osf.io/v5jd3/?view_only=16b91c36be044acfa2a079d3bad85616). Further, this study was run on MTurk instead of Prolific. We conducted an a priori power analysis using the powerplot function of the RRreg package. Given α = .05, the required sample size to detect a correlation of .10 between approval rates and the proportion of dishonest individuals was N = 1,500 (power > .88). As MTurk does not automatically provide the approval rates of the participants, we opened individual batches for each approval rate between >80 and 100. For each approval rate (81–100), we opened a batch for 125 participants. Thus, 20 batches with 125 participants each were opened resulting in a maximum sample size of 2,500 participants. The reason why we thus "oversampled" the suggested 1,500 participants was that, based on current research (e.g., Robinson, Rosenzweig, Moss, & Litman, 2019), we expected that less than 125 participants could be recruited for batches of lower approval rates (i.e., 81–90). Indeed, after 1 week, we had less than 1,500 participants overall (N = 1,027), because there were too few participants in the lower batches. In line with our preregistration, additional batches were opened for very high approval rates (i.e., 98, 99, and 100) until 1,500 participants were reached. We only recruited participants that had more than 100 HITS, as workers with less than 100 HITS always have an approval rate of 100 regardless of how many studies were accepted/rejected.

Participants were relatively heterogeneous with respect to gender (42.33% female, 57.40% male, 0.27% other) and age (M = 33.02, SD = 9.72 years). After consenting to participate in the study, participants provided demographic information. Next, as in Study 1, the participants were asked to participate in an adapted version of the Mind Game paradigm. The participant received a monetary payoff of \$0.40, in addition to their flat fee for participation (\$0.40).

6.1.2 | Results

The same analytical approach as in Studies 1, 2, and 3 was implemented. As in Study 1, the expected percentage of winners was 12.5%. A total of 66.07% of the participants indicated a matching number, which is significantly different from the stochastic baseline of 12.5%, Z = 43.80, p < .001. The proportion of dishonest individuals was estimated to d = .61, SE = .01. The modified logistic regression showed that individuals with lower approval rates were more likely to be dishonest (estimate = −0.18, SE = 0.06, Wald test = 8.89, p = .002, OR = 0.83) as illustrated in Figure 1. Further, another modified logistic regression including age, gender, and approval rate as predictors indicated that age (estimate = −0.66, *SE* = 0.07, Wald test = 80.02, $p < .001$, OR = 0.52) and approval rates (estimate = −0.19, *SE* = 0.07, Wald test = 8.13, $p = .004$, OR = 0.83), but not gender (estimate = 0.21, *SE* = 0.13, Wald test = 2.86, $p = .091$, OR = 1.24), were significant predictors. This indicates that younger individuals with lower approval rates were more likely to be dishonest.

6.1.3 | Discussion

Study 4 replicated the findings of Studies 1–3 on a different crowdworking platform, namely, MTurk.

6.2 | Exploratory analyses across Studies 1–4

Although our hypothesis that lower approval rates are linked to a higher proportion of dishonest individuals was supported, we ran several further exploratory analyses.[5] First, we calculated an additional exploratory modified logistic regression including the quadratic term of the approval rates in Study 4, which was found to describe the data significantly better than the original model (ΔG^2 (1) = 16.00, $p < .001$). Following this exploratory finding, we also tested whether curvilinear models are superior in Studies 1–3. A curvilinear model was found to describe the data better in Study 3 (ΔG^2 (1) = 10.79, $p = .001$) but neither in Study 1 (ΔG^2 (1) = 0.03, $p = .860$) nor Study 2 (ΔG^2 (1) = 0.19, $p = .660$). Plots including the curvilinear models for Studies 3 and 4 can be found in Figure S1, showing inverted U-shaped relations between approval rates and the proportion of dishonest individuals (i.e., there is a lower proportion of dishonest individuals among people with particularly low and high approval rates as compared with people with intermediate approval rates).

As previous research has found that honest and dishonest behavior can be influenced by time restrictions (e.g., Köbis, Verschuere, Bereby-Meyer, Rand, & Shalvi, 2019; Shalvi, Eldar, & Bereby-Meyer, 2012), we further investigated whether time taken for the study impacts proportions of dishonest individuals. Also, we tested whether a measure of experience with online studies—namely, the overall number of submitted studies (as sum score of rejected and accepted studies provided by the export function on Prolific)—had an impact on proportions of dishonest individuals. Specifically, for Studies 1–3, we ran a modified logistic regression including approval rates, time taken, and overall number of studies as predictors of the proportion of dishonest individuals. As MTurk does not provide an exact overall number of studies per participant, we ran a modified logistic regression including approval rates and time taken as predictors of the proportion of dishonest individuals for Study 4. As Prolific provides a time taken variable via the export function and formr provides start and end dates, there were two time variables available for Studies 1–3. The time taken variable from the export function in Prolific captures the time between starting a study on Prolific and entering the study code on the Prolific.co platform. The time taken variable on formr captures when the study was started and when the participants reached the last page (but not when they left the study). Both measures were highly correlated across studies (*rs* > .92) once outliers (i.e., participants that took longer than 30 min to complete the study) were removed. To be completely transparent, we ran separate analyses for including both measures for Studies 1–3. The results were very similar, so that we herein thus report results with the time taken variable from formr. Results including the time taken variable from Prolific can be found in the Supporting Information.

Approval rates were found to be a significant predictor in Studies 1, 3, and 4 (estimates < −0.19, *SEs* < 0.17, Wald tests > 5.37, *ps* < .027, ORs < 0.83) but not in Study 2 (estimate = −0.17, *SE* = 0.09, Wald test = 3.84, $p = .101$, OR = 0.84). Time taken was found to be a significant predictor in Study 2 (estimate = −0.51, *SE* = 0.26, Wald test = 3.93, $p = .022$, OR = 0.60) but not in Studies 1, 3, and 4 (estimates < |0.19|, *SEs* > 0.06, Wald tests < 0.80, *ps* > .373, ORs > 0.83). The overall number of studies was found to be a significant predictor in Study 2 (estimate = −0.51, *SE* = 0.26, Wald test = 3.93, $p = .022$, OR = 0.60) but not in Studies 1 and 3 (estimates < |0.31|, SEs > 0.09, Wald tests < 3.06, *ps* > .111, ORs > 0.73).

Lastly, we conducted an internal meta-analysis across Studies 1–4. Heterogeneity across studies was insignificant as indicated by the Hedges estimator ($Q = 1.40$, $df = 3$, $p = .705$), and the I^2 statistic indicated that 0.00% of the variability was due to heterogeneity rather than chance. A random-effects meta-analysis indicated that the relation between approval rates and proportion of dishonest individuals was significant (OR = .80, 95% CIs [0.74, 0.86], $p < .001$; $k = 4$) as illustrated in Figure 2.

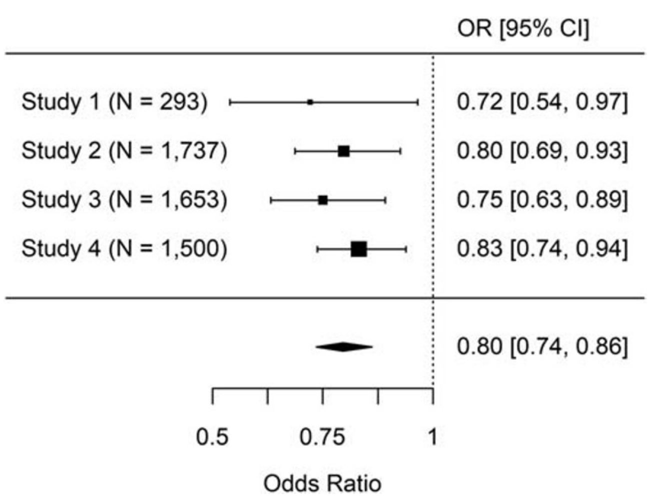

FIGURE 2 Random effects meta-analysis (Forest Plot) of Studies 1–4

[5]We thank the Action Editor as well as the anonymous reviewers for these helpful suggestions.

7 | DISCUSSION

Across four studies and different (versions of) cheating paradigms, crowdworkers' approval rates could be linked successfully to cheating behavior. More precisely, individuals with lower approval rates were more likely to cheat than individuals with higher approval rates in both the Mind Game and the coin flip paradigm (using two different variants of the latter). Importantly, this relation was also found beyond commonly used approval rate thresholds for study inclusion (e.g., a Prolific score of at least 90; Ensor et al., 2019) and on two commonly used crowdworking platforms, namely, Prolific and MTurk. This study thus makes two main contributions: First, it adds to previous studies linking behavior in cheating paradigms to real-world socially questionable behavior (e.g., Cohn et al., 2015; Cohn & Maréchal, 2018; Dai et al., 2017). Second, it identifies approval rates as a characteristic that is systematically linked to dishonest behavior. Given that a large percentage of recent studies on dishonest behavior has been conducted via crowdworking platforms, such as MTurk (e.g., Pfattheicher & Keller, 2018) and Prolific (e.g., Jaffé, Greifeneder, & Reinhard, 2019) and that approval rates or similar metrics are typically available, this might help to increase the interpretability of research findings. More precisely, the results of our investigation strongly suggest to include approval rates (or similar metrics) as another control variable in such platforms. Further, requestors and researchers trying to avoid dishonest participants might use approval rates as a filter.

The underlying idea of this investigation is that people show somewhat similar kind of behavior across situations and is indeed supported by the observed findings; that is, the finding that crowdworkers who are more likely to cheat in a cheating paradigm are also more likely to show more socially questionable behavior in other situations (namely, other tasks on the crowdsourcing platform) is well-aligned with personality trait theory (Allport, 1961), which assumes that individuals do have rather stable personality characteristics that influence behaviors, thoughts, and emotions across different contexts (Zettler et al., 2019). Our study thus not only supports the validity of cheating paradigms in terms of that they can be used in experiments as a (rough) indicator of real-life socially questionable behavior but can also be linked to (meta-analytic) evidence (Zettler et al., 2020) that people with certain traits show a wide range of socially questionable behavior including cheating (in cheating paradigms).

Although Studies 1 and 4 relied on the same cheating paradigm, it should be noted that the proportion of dishonest individuals differed strongly between the studies (d = .23 and d = .61, respectively). This effect might best be explained by the use of two different platforms for the recruitment of participants (Prolific and MTurk, respectively). Indeed, recent studies report higher proportions of dishonest individuals in a single coin flip paradigm on MTurk (d = .48; Pfattheicher & Keller, 2018) than on Prolific (d = .36; Jaffé et al., 2019).[6] In line with this, a recent meta-analysis (Gerlach et al., 2019) showed that participants recruited via MTurk act more dishonest than other populations such as students. Future studies might thus also consider the platform on which the cheating paradigms are run, although one can (so far) only speculate about potential reasons for such observed differences.

Despite the consistency of the findings across Studies 1–4, relations between approval rates and cheating behavior were relatively weak overall. This is likely because approval rates are affected by not only dishonesty itself but also certain other factors such as sloppiness or actual performance; that is, in contrast to other studies linking cheating behavior in paradigms to "pure" real-life dishonesty (e.g., Dai et al., 2017; Kröll & Rustagi, 2016), our outcome measure can only be expected to be partly influenced by dishonesty. In fact, some tasks and studies on crowdworking platforms might even not allow for pure dishonesty. On the other hand, Prolific Team (2018) lists participants' behavior such as "little effort," "failed attention checks," and "lying [the] way into [a] study" as potential reasons for valid rejections, which can—at least partly—be labeled as dishonest or socially questionable behavior. In a similar vein, deception of the requester has also been listed as a valid rejection reason on MTurk (Johnstone et al., 2018). Future studies could set out to test potential explanations for the link between cheating behavior in paradigms and approval rates by testing which kinds of dishonesty affect approval rates.

Further, exploratory analyses suggested that relations between approval rates and dishonest behavior are better described by a curvilinear (namely, an inverted U-shaped) model in Studies 3 and 4. One potential reason for this could be that participants with lower scores try to act more honest in order to increase their scores as they might have noticed not being invited to many tasks/studies anymore. However, another explanation might be that participants with (very) low scores tend to provide random responses in surveys (e.g., Kennedy, Clifford, Burleigh, Jewell, & Waggoner, 2018) and thus might not be actively pursuing the chance to cheat in the cheating paradigm, leading to the inconsistent results concerning the curvilinear pattern across all four studies.

Although our investigation adds to the literature linking behavior in cheating paradigms to real-world behavior, this literature is still comparably small (i.e., including our study, only seven studies linked lab cheating to real-world behavior). Considering that the recent meta-analysis by Gerlach et al. (2019) included 565 experiments, which used lab-cheating paradigms to investigate different aspects of dishonesty, it should thus be a priority of future studies to further investigate how (well) cheating paradigms translate into real-world behaviors.

ACKNOWLEDGEMENTS

This investigation was funded by grants from the Carlsberg Foundation (Carlsbergfondet, CF16-0444) as well as the Independent Research Fund Denmark (Det Frie Forskningsråd, 7024-00057B) to the last author.

[6]Note that d = .36 as reported in Jaffé et al. (2019) is still larger than the proportions of dishonest individuals in Studies 1–3. As suggested by a recent study (Garbarino, Slonim, & Villeval, 2018), this is likely to be explained by relatively low winning probabilities in the mind game (i.e., 16.66%) and the repeated coin flip (i.e., 25%) as compared with a single coin flip (i.e., 50%).

ORCID

Christoph Schild https://orcid.org/0000-0002-6668-5773
Lau Lilleholt https://orcid.org/0000-0002-7257-2636
Ingo Zettler https://orcid.org/0000-0001-6140-7160

REFERENCES

Alexander, L. M., Salum, G. A., Swanson, J. M., & Milham, M. P. (2020). Measuring strengths and weaknesses in dimensional psychiatry. *Journal of Child Psychology and Psychiatry*, *61*, 40–50. https://doi.org/10.1111/jcpp.13104

Allport, G. W. (1961). *Pattern and growth in personality*. Holt: Reinhart & Winston.

Arslan, R. C., & Tata, C. (2019). Chain simple forms/surveys into longer runs using the power of R to generate pretty feedback and complex designs Zenodo https://formr.org. https://doi.org/10.5281/zenodo.3229668

Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, *52*, 376–387. https://doi.org/10.3758/s13428-019-01236-y

Balasubramanian, P., Bennett, V. M., & Pierce, L. (2017). The wages of dishonesty: The supply of cheating under high-powered incentives. *Journal of Economic Behavior & Organization*, *137*, 428–444. https://doi.org/10.1016/j.jebo.2017.03.022

Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73–78. https://doi.org/10.1016/j.joep.2010.12.001

Cohn, A., & Maréchal, M. A. (2018). Laboratory measure of cheating predicts school misconduct. *The Economic Journal*, *128*(615), 2743–2754. https://doi.org/10.1111/ecoj.12572

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies*, *82*(4), 1289–1308. https://doi.org/10.1093/restud/rdv025

Dai, Z., Galeotti, F., & Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, *64*(3), 1081–1100. https://doi.org/10.1287/mnsc.2016.2616

Danske Bank. (2018). Report on the Non-Resident Portfolio at Danske Bank's Estonian branch; available at < danskebank.com/-/media/danske-bank-com/file-cloud/2018/9/report-on-the-non-resident-portfolio-at-danske-banks-estonian-branch-.-la=en.pdf >.

Del Monte, A., & Papagni, E. (2001). Public expenditure, corruption, and economic growth: The case of Italy. *European Journal of Political Economy*, *17*(1), 1–16. https://doi.org/10.1016/S0176-2680(00)00025-2

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory & Cognition*, *47*(1), 182–193. https://doi.org/10.3758/s13421-018-0858-9

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—An experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/jeea.12014

Garbarino, E., Slonim, R., & Villeval, M. C. (2018). Loss aversion and lying behavior. *Journal of Economic Behavior and Organization*, *158*, 379–393. https://doi.org/10.1016/j.jebo.2018.12.008

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*(1), 1–44. https://doi.org/10.1037/bul0000174

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384–394. https://doi.org/10.1257/0002828053828662

Grysman, A. (2015). Collecting narrative data on Amazon's Mechanical Turk. *Applied Cognitive Psychology*, *29*(4), 573–583. https://doi.org/10.1002/acp.3140

Gyimah-Brempong, K. (2002). Corruption, economic growth, and income inequality in Africa. *Economics of Governance*, *3*(3), 183–209. https://doi.org/10.1007/s101010200045

Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, *9*(3), 262–290. https://doi.org/10.1257/pol.20150029

Heck, D. W., & Moshagen, M. (2018). RRreg: An R package for correlation and regression analyses of randomized response data. *Journal of Statistical Software*, *85*(1), 1–29. https://doi.org/10.18637/jss.v085.i02

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*(Supplement C), 72–88. https://doi.org/10.1016/j.jrp.2015.04.003

Hydock, C. (2018). Assessing and overcoming participant dishonesty in online data collection. *Behavior Research Methods*, *50*(4), 1563–1567. https://doi.org/10.3758/s13428-017-0984-5

Ipeirotis, P. G. (2010). Analyzing the Amazon Mechanical Turk Marketplace (SSRN Scholarly Paper ID 1688194). Social Science Research Network. https://papers.ssrn.com/abstract=1688194

Jaffé, M. E., Greifeneder, R., & Reinhard, M. A. (2019). Manipulating the odds: The effects of Machiavellianism and construal level on cheating behavior. *PLoS ONE*, *14*(11), e0224526–e0224526. https://doi.org/10.1371/journal.pone.0224526

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, *93*, 328–336. https://doi.org/10.1016/j.jebo.2013.04.003

Johnstone, D., Tate, M., & Fielt, E. (2018). Taking rejection personally: An ethical analysis of work rejection on Amazon Mechanical Turk [Conference [Discontinued]]. Proceedings of the 26th European Conference on Information Systems (ECIS 2018). https://eprints.qut.edu.au/124689/

Jolley, D., Douglas, K. M., Leite, A. C., & Schrader, T. (2019). Belief in conspiracy theories and intentions to engage in everyday crime. *British Journal of Social Psychology*, *58*(3), 534–549. https://doi.org/10.1111/bjso.12311

Judge, W. Q., McNatt, D. B., & Xu, W. (2011). The antecedents and effects of national corruption: A meta-analysis. *Journal of World Business*, *46*(1), 93–103. https://doi.org/10.1016/j.jwb.2010.05.021

Kennedy, R., Clifford, S., Burleigh, T., Jewell, R., & Waggoner, P. (2018). The Shape of and Solutions to the MTurk Quality Crisis (SSRN Scholarly Paper ID 3272468). Social Science Research Network. https://papers.ssrn.com/abstract=3272468

Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science*, *14*(5), 778–796. https://doi.org/10.1177/1745691619851778

Kocher, M. G., Schudy, S., & Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, *64*(9), 3995–4008. https://doi.org/10.1287/mnsc.2017.2800

Kröll, M., & Rustagi, D. (2016). Shades of dishonesty and cheating in informal milk markets in India (Working Paper No. 134). SAFE Working Paper. https://www.econstor.eu/handle/10419/130760

Lilleholt, L., Schild, C., & Zettler, I. (2020). Not all computerized cheating tasks are equal: A comparison of computerized and non-computerized versions of a cheating task. *Journal of Economic Psychology*, *78*, https://doi.org/10.1016/j.joep.2020.102270

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

Mo, P. H. (2001). Corruption and economic growth. *Journal of Comparative Economics*, *29*(1), 66–79. https://doi.org/10.1006/jcec.2000.1703

Moshagen, M., & Hilbig, B. E. (2017). The statistical analysis of cheating paradigms. *Behavior Research Methods*, *49*(2), 724–732. https://doi.org/10.3758/s13428-016-0729-x

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pfattheicher, S., & Keller, J. (2018). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Social Psychology*, *45*(5), 560–566. https://doi.org/10.1002/ejsp.2122

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, *87*, 26–33. https://doi.org/10.1016/j.euroecorev.2016.03.004

Prolific Team. (2018). Approving submissions and rejecting participants. Prolific. http://researcher-help.prolific.co/hc/en-gb/articles/360009377834-Approving-Submissions-and-Rejecting-Participants

Robinson, J., Rosenzweig, C., Moss, A. J., & LItman, L. (2019). Tapped out or barely tapped? Recommendations for how to Harness the Vast and Largely Unused Potential of the Mechanical Turk Participant Pool. https://doi.org/10.31234/osf.io/jq589

Schild, C., Heck, D. W., Ścigała, K. A., & Zettler, I. (2019). Revisiting REVISE: (Re)testing unique and combined effects of REminding, VIsibility, and SElf-engagement manipulations on cheating behavior. *Journal of Economic Psychology*, *75*, https://doi.org/10.1016/j.joep.2019.04.001

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, *23*(10), 1264–1270. https://doi.org/10.1177/0956797612443835

Teubner, T., Hawlitschek, F., & Adam, M. T. P. (2019). Reputation transfer. *Business & Information Systems Engineering*, *61*(2), 229–235. https://doi.org/10.1007/s12599-018-00574-z

Thelin, J. R. (2019). An embarrassment of riches: Admission and ambition in American Higher Education. *Society*, *56*(4), 329–334. https://doi.org/10.1007/s12115-019-00376-3

Todd, J., Aspell, J. E., Barron, D., & Swami, V. (2019). Multiple dimensions of interoceptive awareness are associated with facets of body image in British adults. *Body Image*, *29*, 6–16. https://doi.org/10.1016/j.bodyim.2019.02.003

Zettler, I., Thielmann, I., Hilbig, B., & Moshagen, M. (2019). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *erspectives on Psychological Science*. https://doi.org/10.1177/1745691619895036

Volkswagen. (2015). Annual Report 2015; available at < annualreport2015.volkswagenag.com/ >.

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 Difference*, *81*, 107–111. https://doi.org/10.1016/j.paid.2014.10.007

AUTHOR BIOGRAPHIES

Christoph Schild postdoctoral researcher at the Department of Psychology, University of Siegen (Germany)

Lau Lilleholt is a PhD student at the Department of Psychology, University of Copenhagen (Denmark).

Ingo Zettler is Professor with Special Responsibilities in Personality and Social Behavior at the Department of Psychology, University of Copenhagen (Denmark).

How to cite this article: Schild C, Lilleholt L, Zettler I. Behavior in cheating paradigms is linked to overall approval rates of crowdworkers. *J Behav Dec Making*. 2020;1–10. https://doi.org/10.1002/bdm.2195