*Cheaters, Liars, or Both? A New Classification of Dishonesty Profiles*



**As Published:** 10.1177/0956797620929634

**Publisher:** SAGE Publications

**Persistent URL:** https://hdl.handle.net/1721.1/133974

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

# Cheaters, liars, or both?
# A new classification of dishonesty profiles


David Pascual-Ezama[1,2], Drazen Prelec[2], Adrián Muñoz[3] & Beatriz Gil-Gómez de Liaño[3,4,5]

[1] Universidad Complutense de Madrid

[2] Sloan School of Management, MIT

[3] Universidad Autónoma de Madrid

[4] Brigham & Women's Hospital-Harvard Medical School

[5] University of Cambridge


RUNNING HEAD: Dishonesty & Moral Decisions

Authors Note:


David Pascual-Ezama, Universidad Complutense de Madrid & MIT Sloan School of Management Fulbright Fellow; Drazen Prelec, Sloan School of Management, Department of Economics, Department of Brain and Cognitive Sciences, MIT; Adrian Muñoz, Universidad Autónoma de Madrid; Beatriz Gil-Gómez de Liaño, Visual Attention Lab at Brigham and Women's Hospital- Harvard Medical School, University of Cambridge & Departamento de Psicología Social y Metodología, Universidad Autónoma de Madrid.

Correspondence concerning this article should be addressed to David Pascual-Ezama, Accounting and Finance department, Campus de Cantoblanco s/n, 28223, Madrid, Spain.

Email: dp_ezama@mit.edu; david.pascual@ccee.ucm.es.

**Abstract**

Experimental studies of dishonesty usually rely on population-level analyses, which compare the distribution of claimed rewards in an unsupervised, self-administered lottery (e.g., tossing a coin) against the expected lottery statistics (e.g., 50%-50%). Here we provide a paradigm that measures dishonesty at the individual level, and identifies new dishonesty profiles with specific theoretical interpretations. We find that among dishonest participants, (a) some do not bother implementing the lottery at all, (b) some implement but lie about the lottery outcome, and (c) some violate instructions by repeating the lottery multiple times until obtaining an "acceptable" outcome. These results hold both in the lab and with online (Mturk) participants. In Experiment 1 (N=178), the lottery was a coin toss, permitting only a binary honest-dishonest response; Experiment 2 (N=172) employed a 6-sided die roll, which permitted gradations in dishonesty. We replicate some previous results, and also provide a new, richer classification of dishonest behavior.

*Keywords: Dishonesty, Moral decision-making, Behavioral profiles.*

Abstract word count: 150 words.

Authors Contributions:

D. Pascual-Ezama developed the study concept. All authors contributed to the study design. Testing and data collection were performed by D. Pascual-Ezama, A. Muñoz and B. Gil-Gómez de Liaño. All authors contributed to the data analysis and interpretation. D. Pascual-Ezama drafted the manuscript and D. Prelec and B. Gil-Gómez de Liaño provided critical revisions. All authors approved the final version of the manuscript for submission.

Ethical Considerations:

This study has been approved by the Universidad Autónoma de Madrid IRB.

**Acknowledge**

This study was made possible thanks to the funding received by the Fulbright Commission award FMECD-ST-2017 to David Pascual-Ezama.

## Introduction

Dishonesty imposes massive costs on organizations and society, as one can see from striking cases of institutional and corporate corruption that confront us on a daily basis. Detecting and discouraging dishonest behavior is a major task facing both the public and private sectors. Understanding the mechanisms behind dishonesty is likewise a fundamental research challenge for psychology and other social sciences (Mazar & Ariely, 2006; Ayal et al., 2015). Psychological (e.g., Monin & Miller, 2001), sociological (e.g., Tembrunsel & Messick, 2004), and philosophical (e.g., Finch, 2011) research has provided some evidence and insights into variables that modulate dishonest behavior. For example, it is known that dishonesty decreases with the probability of detection, and that self-reported measures cannot be fully trusted.

For obvious reasons, it is not straightforward to elicit and observe dishonesty from participants in the experimental lab. People may refrain from dishonest behavior if they perceive that their dishonesty might be detected — a reasonable assumption in many laboratory settings. Much effort has been devoted in the last few years in developing experimental tasks that accurately detect underlying dis(honest) preferences. These can be divided into several categories, specifically: Population inferred cheating tasks, individually inferred cheating tasks, social tasks, and field tasks (see Jacobsen, Fosgaard & Pascual-Ezama, 2018, for a review).

By ‘population inferred tasks’ we refer to tasks where cheating can only be observed at the aggregate, statistical level. This requires a known statistical distribution of expected outcomes or an equivalent control group, whose performance is known. The methodological advantage of these tasks is that they give respondents an opportunity to deceive with impunity. Examples of population inferred cheating tasks with a known expected outcome distribution include rolling a die (or multiple dice) or flipping a coin in private. In these tasks, participants receive (or do not receive) a reward depending on the result (Bucciol & Piovesan, 2011; Fischbacher & Follmi-Heusi, 2013; Belot & Schroder, 2015). Here, only the participant knows the actual outcome and is therefore sure that the experimenter cannot detect them if they act dishonestly. The experimenter can only infer statistically that some proportion of participants have cheated. In spite of this limitation, inferred tasks do provide information about cheating in unsupervised situations. In particular, one can infer the proportion of people that are honest, that are dishonest, and that are simply lucky in obtaining the maximum reward. Interestingly, dishonest individuals can be further divided into subtypes according to how extreme their dishonesty is. For instance, if rolling the die and getting a six (the non-reward bad outcome) in a classic task (Fischbacher & Follmi-Heusi, 2013), some may decide to report five (getting the highest reward), but they can also report three or four, perhaps to adjust the response up to their ethical 'thresholds', or possibly because they think that a lower claimed reward is somehow less likely to be detected.

Unfortunately, inferred tasks do not identify dishonesty at the individual level, and do not discriminate between different modes of participant misbehavior, namely, between cheating (not following experimental instructions in some way) and explicit lying (misreporting an event such as the result of a coin toss).

The studies described below employed an innovative virtual adaptation paradigm based on using participants' own devices. The paradigm allowed detection of individual behavior in any inferred task, while preserving the participants' belief that their behavior cannot be detected, that is, that they could deceive with impunity. We used a modification of the two classic inferred cheating tasks, the die-under-the-cup, and the coin flip. Note that Peer, Acquisti, and Shalvi (2014) used a similar methodology using lab devices with the coin flip task. However, although there are previous virtual adaptations of the die-under-the-cup task, some studies used random external websites (Kobis et al., 2019), and others did not guarantee the perception of impunity (Kocher et al., 2018). To our knowledge, our experiment is the first one to use a die-under-the-cup virtual adaptation task with adult subjects that allows us to register the results of the actual cup toss at the individual participant level (see Markiewicz & Gawryluk, 2019, for a lab study with children). These actual data can then be compared with the reported information at the individual participant level. At the same time, each participant believes that no one (besides themselves) can see the outcome, because they are using their own devices and a website of their choice. Therefore, the participants behave as they like, and can cheat and claim higher outcome without any detection risk.

Like the die-under-the-cup task, a simple coin toss (Bucciol & Piovesan, 2011) also allows the experimenter to statistically determine the data distribution with a 50/50 distribution of the two possible outcomes. However, a drawback of the coin toss is that cheating is dichotomous. With one-coin flip, the participant has a binary choice, whether to be honest or not. Unlike the die-in-a-cup task, here it is not possible to adjust the 'cheating level'. Thus, the two tasks — die-under-the-cup and coin toss — might measure different types of dishonesty, and reveal different cheating profiles.

Going beyond replication of earlier results, our main goal and contribution is to detect precisely *how* individual participants choose to implement their decision to behave dishonestly. As noted already, this is accomplished while completely preserving participants' perception of impunity. Our results allow us to create a new classification of dishonest behaviors based on a distinction between cheating and lying.

## Method & Results

### Experiment 1: Black & White (Dis)Honesty

**Materials and Procedure**.

The main objective of experiment 1 is to provide a fine-grain record of individual behavior in a situation where decisions are entirely private (unobservable) and honesty is potentially costly. To do this, we developed a virtual "black and white" adaptation of the coin toss task (Bucciol & Piovesan, 2011). This task does not allow for gradations in dishonesty, as participants who observed an unlucky black outcome had only two response options: Being honest and reporting black, or being dishonest by reporting white and receiving a substantial additional money payment.

The website www.rollanflip.com is a website created by the researchers in order to register the real result of the task, with the versions 'flip the coin' or 'roll the die' (see Experiment 2 below). Through this website, we can record the result and time so we can match those data with the reported results and time of the participants in the computer report, both at the lab and using Mturk.

Lab participants arrived at the booth laboratory, where there were eight different isolated cabins, all of them closed and with a computer to perform part of the task of the study. Before entering each cabin, every participant signed an informed consent form. They were asked whether they have charged the battery in their cell phones, as they would need them for the study. The experimenter ran the program using E-prime 3.0 (Psychology Software Tools, Pittsburgh, PA) for every participant. They received the same message with instructions three minutes apart to make sure the exact time everyone started the task. The task was to flip a coin using the website www.rollandflip.com, or a similar one. Participants had the option to choose any similar website. If they got 'white', they would get the extra $5 bonus, but if participants got 'black' they would only get the $1 show-up fee. Importantly, to maintain privacy and the perception of impunity in their toss, they were asked to use their cell phones to flip the coin in the website. Participants had to report the result in the computer and receive the money according to the declared outcome. The payment was placed in an envelope inside a box in the cabin. They received $6 if the reported result was white, and only $1 if they reported black. Finally, every participant was assigned a code. We registered the exact time of the participation for every code, in order to subsequently correlate the response reported via the computer with the actual outcome registered in the website. MTurk participants received the same instructions. The difference was that MTurk participants ran the experiment using the MTurk platform out of the lab. Therefore, they were paid according to their report in the platform.

**Participants**.

According to García-García, Reding-Bernal, and López-Alvarenga (2013), the minimum sample for a 2-proportion comparison setting $\alpha$=.05, and power = .95 is about 45 participants per cell. In the present coin-task, and with a probability of p= .50 of lucky participants, and in a within-subjects study we would need about 90 participants. To avoid problems of participants not showing up, we admitted participants in the study for about 100. In the lab, there were 109 registered when we closed the application, and there were 97 in Amazon Mechanical Turk (MTurk). All of them got \$1 as a show-up (participation) fee and the opportunity to earn a \$5 performance-based bonus.

Individuals giving an immediate response (less than 5 seconds after receiving computerized instructions) without using the www.rollandflip.com website were classified as "Radical Dishonest" since they claimed the high money outcome without flipping a coin/roll a die in some other website (5 seconds is insufficient time to go to an alternative website and initiate a coin toss, i.e., a die roll). Individuals who gave a report that took more than 5 seconds and also did not go to our website were eliminated from the final sample. For these individuals, we could not rule out the possibility that they might have used another website (which the instructions permitted), or perhaps, for Mturk participants, tossed a physical coin, which might match the spirit if not the letter of the instructions. Therefore, for these participants we could not unambiguously assess their level of honesty and consistency with instructions. The above exclusion criterion eliminated 16 participants in the lab study, and 12 participants in the Mturk study. The final sample comprised 178 individuals, 93 from the lab (35% male, 65% female; mean age = 19.86, SD = 2.25) and 85 from MTurk (40% male, 60% female; mean age = 36.84, SD = 12.08).

**Results**.

**Population level analysis**. We examined whether the outcome reported differed from the expected proportion by chance, as in classic inferred cheating task analyses. In the Lab treatment, 72 white flips were reported from a total number of 93 flips (77.4%). One side binomial test revealed that the percentage was significantly higher than what would be expected at random ($p= < .01$). Regarding the Mturk sample, 68 white flips were reported from a total number of 85 flips (80%). The same binomial test indicated that the proportion reported was also significantly higher than the theoretically expected proportion ($p=<.01$). Replicating prior results (e.g., Bucciol & Piovesan, 2011; Pascual-Ezama et al., 2015), the data confirm dishonest behavior by some participants at least. Comparing Lab and Mturk samples, chi-square for independence revealed that the two proportions are not significantly different ($\chi^2(1,N=178)=0.176$, $p=0.675$; $\varphi=0.031$), as we can see in Figure 1.

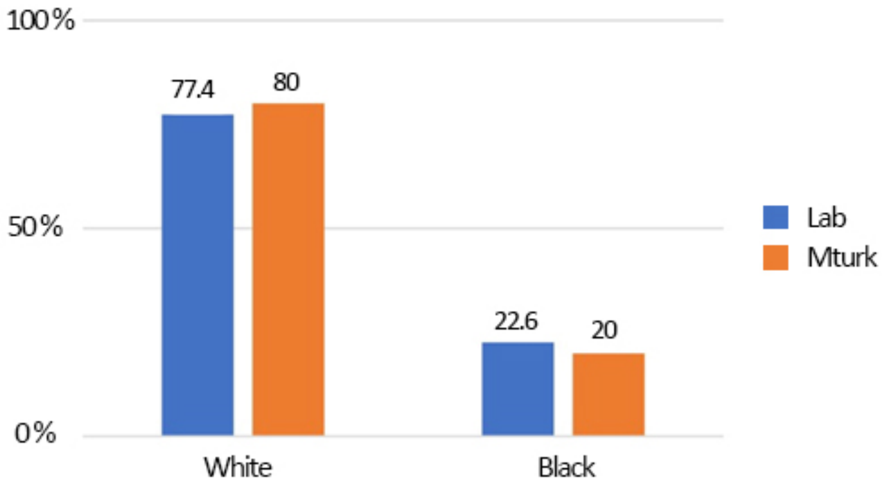


Figure 1. Percentage of coins outcome reported both for Lab and virtual M-Turk samples.

**Individual level analysis**. Our paradigm allows us to match results of the reported outcome in the computer with the real result generated by the website. On this basis we can classify participants in different categories (see Table 1): Lucky participants (those who really got white) comprised 44% and 48% in the university and MTurk samples, respectively; honest participants getting black and reporting black comprised 22% and 20% respectively; dishonest participants (34% and 32%, respectively) who reported white when they really got black.

Because the paradigm lets us discriminate between lucky and dishonest participants, we can obtain a more realistic percentage of (dis)honest people than the one obtained with aggregated, population-level information. If we remove the lucky sample from calculations and take into account only those getting black, the percentage of honest people was 41% in the lab and 37% in Mturk (white columns in Table 1).

Table 1. Black & White (Dis)honesty. Classification of participants according to their reported / actual results.

| | | Lab | | MTurk | |
|---|---|---|---|---|---|
| | | (n=93) | (n=52) | (n=85) | (n=43) |
| Flip the coin - obtain white - report white | LUCKY | 44% | - | 49% | - |
| Flip the coin - obtain black - report black | HONEST | 22% | 41% | 19% | 37% |
| Flip the coin - obtain black - repeat until white - report white | CHEATERS NON-LIARS | 10% | 17% | 3.5% | 7% |
| Flip the coin - obtain black - report white | LIARS | 13% | 23% | 3.5% | 7% |
| Do not flip the coin at all - report white | RADICAL DISHONEST | 11% | 19% | 25% | 49% |

*First grey row show percentage results including ‘lucky’ people. White columns show percentages of total sample excluding ‘lucky’ people.

More importantly, individual-level data yields a finer classification of dishonest behaviors. Specifically, we can divide dishonest participants into at least three clear different profiles — cheaters non-liars, liars and ‘radical dishonest’ (i.e., both cheating and lying). Conceptually, a *cheater* is someone who breaks the rules, while a *liar* is someone who does not tell the truth. As we can see in Table 1, there are people who cheated but did not lie. They flipped the coin, obtained black, but continued flipping the coin several times until they got white, and then they reported white. The instructions were unequivocal regarding the number of flips, only one. That is, they broke the rules of the task, but technically, they did not lie because at some point they received white. We have called these participants ‘cheaters non-liars’, as shown in Table 1. But the data also reveals a substantial proportion of genuine ‘liars’, who flipped the coin, obtained black and reported white. They did not break the rules, they flipped the coin only once as demanded, but as the result implied no reward they lied to claim reward. Finally, we have found an unexpected and surprising new profile, the participants who did not bother even flipping the coin. We refer to them as ‘radical dishonest’ because they both cheated (did not flip the coin) and lied (that the coin came up white). By their immediate response, we can infer they did not go to any website. So, they broke the rules as they did not follow the instructions, and they also lied reporting ‘white’ when they did not even flip the coin.

The above profiles were present both in the lab and MTurk samples. Comparing the results of the two samples, we found no differences between lucky people ($\chi^2$(1,N=178)=0.308, p=0.579; φ=0.042) and honest ones ($\chi^2$(1,N=178)=0.176, p=0.675; φ=0.031). However, there are significant differences in the prevalence of the types of dishonest people between the lab and Mturk samples. Mturk participants were more likely to be radically dishonest that lab participants ($\chi^2$(1,N=178)=6.01, p=0.014; φ=0.184). ‘Cheaters non-liars’ ($\chi^2$(1,N=178)=2.67, p=0.102; φ=0.122) and “liars” ($\chi^2$(1,N=178)=5.06, p=0.025; φ=0.17), were significantly more prevalent in the lab. Maybe, as the general dishonest people behave dishonestly in any case (no differences show up between lab and MTurk samples when combining all dishonest profiles), increased impunity by a virtual application like the MTurk, might people make more radically dishonest than in the ‘in-person’ lab situation.

Regardless of the proportion of individuals in the different profiles, we have found the same four different profiles, both in the lab and Mturk samples: (1) Honest people; (2)

Liars, or what Fischbacher and Follmi-Heusi (2013) called ‘income maximizing subjects’ (note that there is no way to test ‘partial liars’ by using a Black & White paradigm, as Fischbacher and Follmi-Heusi also found); (3) the new ‘radical dishonest’ profile; and (4) ‘cheaters non-liars’, a new profile not previously described in the literature where participants repeated the task cheating to get the desired result to obtain the reward, but technically did not lie.

**Experiment 2: Grey Scale (Dis)Honesty**

Our aim in Experiment 2 is to study intermediate levels of (dis)honest behavior by using a different classical dishonesty task, the die-under-the-cup paradigm. We also wanted to check if the results of Experiment 1 replicate in terms of generating similar diversity of dishonesty profiles.

**Materials and Procedure**.

Experiment 2 used a digital adaptation of the die-under-the-cup task. This task allows us to elicit different levels of (dis)honest behavior as participants not only decide to lie or cheat but also modulate the level of dishonesty. Under this paradigm dishonest participants must decide whether to go (or not) for the maximal reward.

All materials and procedure were the same as in Experiment 1, but the instructions changed as the new participants were asked to roll a die instead of flipping a coin. As in Experiment 1, they were asked to use their cell phone to the die via www.rollandflip.com or a similar website. They were then asked to report the outcome in the computer. Following Fischbacher and Follmi-Heusi’s (2013) rewards system, participants understood that they would receive the dollar equivalent of the number they reported except for the 'unlucky outcome' six, which would result in zero reward. Therefore, by reporting five, they would receive the maximum reward ($5). In this new task participants could choose not only to be (dis)honest, but also in case of dishonesty, adapt their decisions to different levels from maximum to minimum reward. This is particularly important because as we have previously mentioned, Fischbacher and Follmi-Heusi (2013) found 'partial liars' that adapted their level of dishonesty. Under this task we expect to find similar 'partial dishonesty' profiles. Therefore, we will call this decision-making 'Grey Scale Dishonesty'.

MTurk participants received the same instructions as participants in the lab. The only difference was that MTurk participants run the experiment by MTurk platform out of the lab, and they were paid according to their report in the Platform, like in experiment 1.

**Participants**.

To facilitate comparison to experiment 1, we maintained a sample of about 100 people in both studies. A total of 194 new participants agreed to participate in the experiment. They were 95 university students in the lab study, and 99 individuals recruited by MTurk. As in experiment 1, they participated in the study for a $1 show-up fee and the opportunity to earn up to $5 performance-based bonus. For the same reasons as in experiment 1, we eliminate 14 participants from the lab, and 8 from the MTurk samples who responded under 5 seconds after receiving instructions and did not go to our website. The final sample was composed by 172 individuals, 81 university students in the lab (10% male, 90% female; mean age = 19.48, SD = 1.58), and 91 participants in MTurk (47% male, 53% female; mean age = 36.01, SD = 10.70).

**Results**.

**Population-level analysis.** We examined whether the reported outcome distribution differed from a uniform distribution, as in classical inferred tasks aggregated analyses. A Kolmogorov-Smirnov test for one sample showed that both sample distributions were significantly different from the expected uniform distribution ($Z_{Lab}$=1.533, p=0.018, d=0.43, 95 % CI [0.38-0.48]; $Z_{Mturk}$=2.684, p<0.001, d=0.78, 95 % CI [0.73-0.83]), showing that there were people in both samples not telling the truth, that is, the real outcome. Then, we tested for each die outcome, whether its proportion differed from the expected by chance. As we can see in Figure 2, high reward proportions were significantly higher than expected by chance (for statistics see Table 2), while low reward proportions and no reward are instead marginally more moderate than expected. Both in the lab and Mturk, outcome 'five' is significantly above the expected 16.7%, while 'six' is significantly lower than expected. In MTurk, 'one' and 'two' are also considerably lower than expected. That distribution implies that some subjects tended to report more or less than they probably really obtained. According to the results, we also can assume that some people are honest at least. Likewise, and as expected, not all the lying participants 'maximally' lie. Some subjects neither reported the truth nor reported 'five' but, essentially, they reported the outcome 'four'. Intriguingly, this effect is observed mainly in Mturk, as we can also see when comparing both samples (see again Table 2 for statistics).

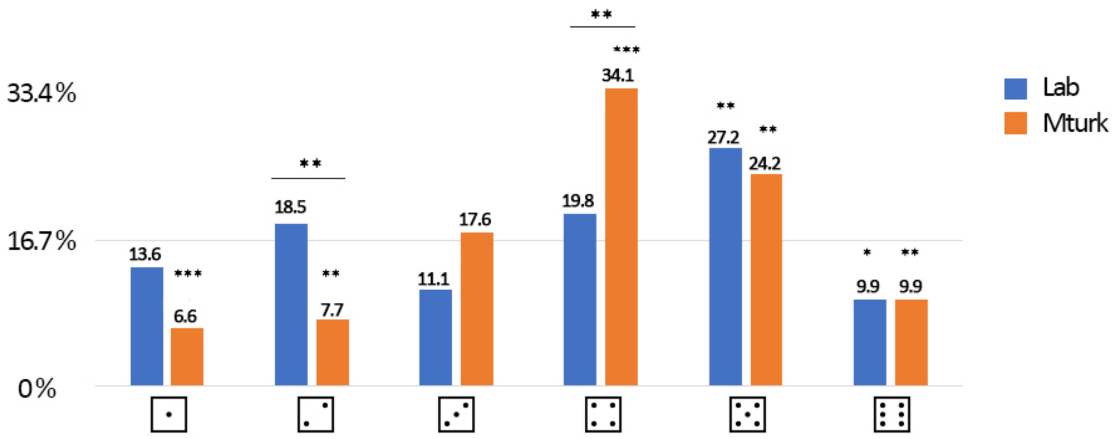


Figure 2. Declared die outcome (Lab vs Mturk). Asterisks above the numbers mean significant differences between the observed and expected distribution by chance. Asterisks over the black lines mean significant differences between lab and Mturk samples. *** p<0.01; ** p<0.05; * p<0.1

Table 2. Binomial test ($\pi$) comparing actual outcomes with uniform distribution. Also, the Chi-squared test ($\chi 2$) comparing the differences between Lab and Mturk samples for each outcome.

| | $\pi$ | | | | |
|---|---|---|---|---|---|
| **Dice** | **Lab** | **Mturk** | $\chi^2$ | $p$ | φ |
| **1** | 0.28 | <0.001 | 2.34 | 0.13 | 0.12 |
| **2** | 0.72 | 0.01 | 4.5 | 0.034 | 0.16 |
| **3** | 0.11 | 0.45 | 1.45 | 0.23 | 0.09 |
| **4** | 0.27 | <0.001 | 4.42 | 0.01 | 0.16 |
| **5** | 0.01 | 0.03 | 0.2 | 0.65 | 0.03 |
| **6** | 0.08 | 0.04 | 0.001 | 0.99 | <0.01 |

**Individual-level analysis.** The population-level analysis indicates dishonesty behavior, but does not discriminate among different dishonest profiles. Individual level analyses provide a more fine-grained picture of different forms of dishonesty that the paradigm allows (see Table 3). If we exclude lucky participants who received the five-outcome (6% in the lab and 11% in the Mturk), we find 64% of honest people in the lab and 57% in Mturk. The profiles of the remaining dishonest participants are similar to those found for the coin task, but the dice paradigm allows us to diversify further the classification. As evident in Table 3, we find that there are people maximizing dishonesty (getting the maximum reward); there are also people who cheat, lie or both, without claiming the maximum reward (which of course they could have done as there was no supervision). As in Experiment 1, the lab and in Mturk studies yield similar profiles (except for "non-maximizing dishonest" profile that does not show up in the lab but in the MTurk sample). Once again, the percentage of 'radical dishonest' people is higher in Mturk ($\chi^2$(1,N=25)=7.719, p=0.005; φ=0.21).

Table 3. Grey Scale (Dis)honesty. Classification of participants according to their reported / actual results.

| | | | Lab (n=81) | Lab (n=76) | MTurk (n=91) | MTurk (n=80) |
|---|---|---|---|---|---|---|
| Roll the die - obtain 5 - report 5 | | LUCKY | 6% | - | 11% | - |
| Roll the die - obtain x - report x | Lucky honest | HONEST | 49% | 53% | 42% | 47% |
| Roll the die - obtain 6 - report 6 | Unlucky Honest | HONEST | 10% | 10.5% | 9% | 10% |
| Roll the die - obtain 6 - repeat until different than 6 - obtain x - report x | Submaximizing Cheaters non-liars | CHEATERS NON-LIARS | 4% | 4% | 2% | 2.5% |
| Roll the die - obtain x different than 5 - repeat until 5 - report 5 | Maximizing cheaters non-liars | CHEATERS NON-LIARS | 10% | 10.5% | 4% | 5% |
| Roll the die - obtain x - report >x but less than 5 | Submaximizing Liars | LIARS | 10% | 10.5% | 9% | 10% |
| Roll the die - obtain x different than 5 - report 5 | Maximizing Liars | LIARS | 5% | 5% | 2% | 2.5% |
| Do not roll the die at all - report <5 | Submaximizing radical dishonest | RADICAL DISHONEST | - | - | 12% | 13% |
| Do not roll the die at all - report 5 | Maximizing radical dishonest | RADICAL DISHONEST | 6% | 6.5% | 9% | 10% |

*Again, grey rows show percentage results including "Lucky" people. White rows show percentages of total sample excluding "Lucky" people.

It is also interesting to compare the real die roll distribution and the reported one. For this purpose, we excluded from the analysis every participant who rolled the die more than once or did not roll the die at all. Differences between both distributions were significant ($Z_{Wilcoxon}$=-2.27, p=0.02, d=0.13, 95 % CI=[0.1,0.15]). Figure 3 shows the proportion for each outcome with the significance of the Wilcoxon test between outcome pairs (reported and really obtained; also see the statistics in Table 4). These results show several interesting effects. Firstly, we can observe how outcome 'one' proportion differs significantly from the real 'one' proportion, which is more evident for the lab sample. Secondly, in general, outcomes 'four' and 'five' are clearly over-reported, while 'one' and 'six' real outcomes are under-reported. Finally, the main difference between lab and Mturk samples is that for outcomes 'one' ($\chi^2$(1,N=130)=1.88;p<0.08;φ=0.11) and 'two' ($\chi^2$(1,N=130)=4.9;p<0.01;φ=0.2) people report significantly less in the Mturk than in the lab while in outcome "four" we found the opposite ($\chi^2$(1,N=130)=3.84;p<0.02;φ=0.17). Again, as in Experiment 1, it seems that increased impunity in the MTurk paradigm might make people more radically dishonest. It may worth mentioning too that the probability of cheating is a function of your dice roll (so, if you get a six you are more likely to cheat) especially with "MTurkers" (F(1,89)=58.31,p<0.001;$\eta^2$=0.39, 95% CI=[0.24-0.52]). This seems to be also a function of age in MTurk sample, older people cheat less (F(1,89)=4.63, p=0.034;$\eta^2$=0.05, 95% CI=[0.01-0.16]).

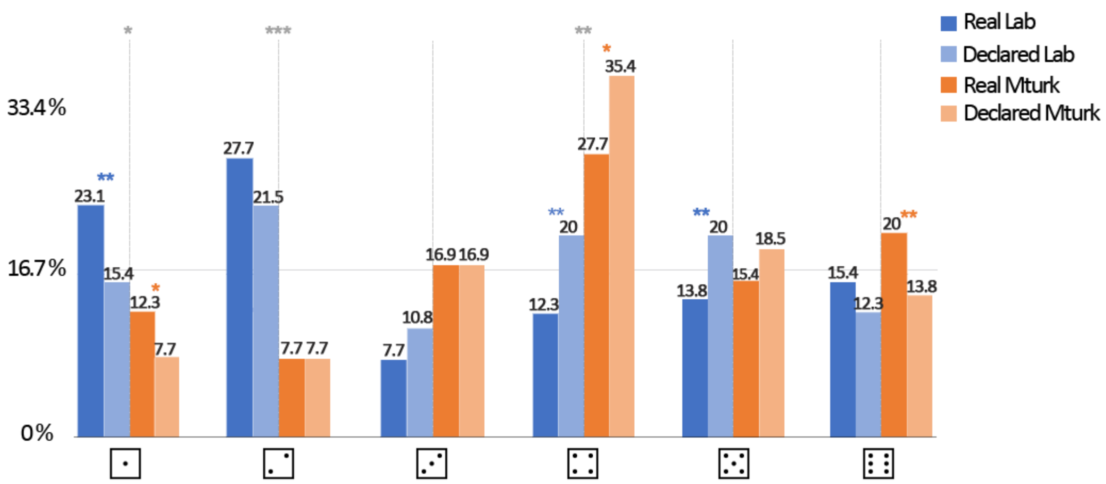


Figure 3. Declared vs. Real die outcome (Lab vs. Mturk). Asterisk above the bars mean significant differences between the declared and real outcomes. Asterisks over the black dashed lines mean significant differences between lab and Mturk samples. *** p<0.01; ** p<0.05; * p<0.1

Table 4. Wilcoxon tests comparing real and declared outcomes for both Lab and Mturk samples. Statistics (z), significance (p), effect sizes (Cohen-d) and Confidence intervals (lower-LCI and upper-UCI).

| | **Dice** | ***z*** | ***p*** | ***Cohen-d*** | ***LCI*** | ***UCI*** |
|---|---|---|---|---|---|---|
| *Lab* | **1** | 2.23 | 0.03 | 0.17 | 0,00 | 0.34 |
| | **2** | 1.63 | 0.1 | 0.11 | -0.06 | 0.28 |
| | **3** | 1.41 | 0.16 | -0.14 | -0.3 | 0.03 |
| | **4** | 2.23 | 0.03 | -0.2 | -0.37 | -0.03 |
| | **5** | 2,00 | 0.04 | -0.16 | -0.33 | 0.01 |
| | **6** | 1.41 | 0.15 | 0.14 | -0.03 | 0.31 |
| *Mturk* | | | | | | |
| | **1** | 1.73 | 0.08 | 0.16 | -0.01 | 0.33 |
| | **2** | 0.01 | 0.99 | 0,00 | -0.17 | 0.17 |
| | **3** | 0.01 | 0.99 | 0,00 | -0.17 | 0.17 |
| | **4** | 2.21 | 0.03 | -0.15 | -0.32 | 0.02 |
| | **5** | 1.41 | 0.16 | -0.07 | -0.24 | 0.1 |
| | **6** | 2,00 | 0.04 | 0.15 | -0.02 | 0.32 |

## Discussion & Conclusions

The results presented here provide a new, theoretically suggestive classification of dishonest behavior. Recall that in our paradigm participants face three implicit decision points, each providing a different opportunity to deceive. The first decision is whether to toss a coin / roll a die at all. If they do comply at this point but then receive a suboptimal outcome, they then face a second decision, whether to repeat the toss/die roll, thereby violating explicit instructions to toss/die roll only once. If they do not repeat, they then have to decide whether to report honestly or to lie (and with the die roll, how much to lie).

This sequence of decisions generates a spectrum of dishonesty profiles — of cheating without lying, lying without cheating, and both cheating and lying. We see in experiment 2 that people also modulate the level of dishonesty, replicating results from previous studies (Fischbacher & Follmi-Heusi, 2013). We have therefore an experimental paradigm that elicits distinct dishonesty profiles unobtrusively at the individual level, and that can be combined in future studies with additional manipulations and variables. Such a fine-grained picture could not be obtained from aggregated data with population-level analyses.

The results obtained so far already provide a richer picture of individual dishonesty than available previously. First, we confirm the presence of 'completely' honest people, in line with previous results, e.g., Fischbacher and Follmi-Heusi (2013). Second, we find a substantial proportion of people that flip the coin until they obtain white or roll the die until they obtain an "acceptably" high payoff result. This fits with prior work on self-serving justifications. Shalvi et al. (2011) found that when people are allowed to roll the die more than once, but only the first roll was valid for reporting purposes, the highest outcome was sometimes reported (even if it was not the first one). Of course, participants are always able to flip the coin or roll the die more than once (similar results with children were found in Maggian & Villeval, 2016).

Why might people engage in what seems to be a ritualistic exercise, to flip the coin multiple times? One can interpret this behavior from the standpoint of self-signaling (Bodner and Prelec, 2002; Prelec and Bodner 2003). Reporting an untrue outcome is a signal of a general character trait, namely, that one is willing to state something that is false for a monetary benefit. Put another way, lying participants are exposed (to themselves) as people whose false testimony can be bought for a relatively paltry sum of money. This may have implications about self-image beyond the narrow scope of the experiment. In contrast, cheating participants who toss multiple times are only failing to follow instructions to the letter. The negative self-image implications of violating the 'labor contract' between them and the experimenter is more narrow in this case.

Among the people who roll more than once, some roll die until they get the maximum reward ($5), while others stop at a smaller payoff outcome. These 'submaximizing cheaters non-liars' are willing to break the no-repetition rule in principle, but not to ignore it altogether. The fact that they are not claiming the maximum reward outcome may provide some protection to their self-esteem.

Interestingly, we see in our data a new deception profile of participants who do not flip the coin or roll a die at all (they are more numerous in the Mturk sample). This profile may also be present in 'traditional' aggregate data results, but can only be detected with an individual-level paradigm like ours.

Given that some participants take this radical shortcut to maximum reward, one can ask why would *any* lying participant bother flipping the coin / rolling the die at all? If they knew in advance that they would lie about an unfavorable outcome, what is achieved by tossing a coin, whose outcome will be ignored? There are two possible interpretations of tossing and then lying. One possibility is that even though a participant knows that they will lie if the outcome is bad, the psychological cost to their self-image is only absorbed with an actual lie, not with a hypothetical lie. By tossing the coin they may be hoping for 'moral luck' (Williams, 1981), i.e., obtaining the maximum reward 'without having to lie.' A different possibility is self-deceptive prior beliefs (Mijovic-Prelec and Prelec, 2010): Before tossing the coin, participants may believe that they will report honestly, as this belief is pleasant to their self-image, but then change their mind when the outcome is bad.

Comparing the results from experiments 1 and 2 we find that the 'Grey Scale Dishonesty' with the die roll task generates a higher proportion of honest responses. It could be that the cost of honesty is less with the die roll, as "honest" participants still receive some additional compensation, while being able to 'proudly' affirm their honesty.

It is important that the same set of profiles is elicited with different protocols (coin toss and die roll) and in different experimental settings and participant populations (lab students and Mturk workers). While the exact proportions of profiles are slightly different depending on tasks and population, the underlying consistency in types suggests that we are dealing with a relatively robust classification that has relevance outside of the artificial setting of a psychological experiment. Organizations concerned about dishonesty when making their selection, promotion, and compensation decision, should appreciate that honesty is not a one-dimensional trait, and that dishonest behavior is often self-contradictory, with individuals breaking rules and acknowledging them at the same time.

## References


Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three Principles to REVISE People's Unethical Behavior. *Perspectives on Psychological Science*, 10(6), 738-741. doi: 10.1177/1745691615598512

Belot, M., & Schroder, M. (2015). The spillover effects of monitoring: a field experiment. *Management Science*, 62(1), 37-45. doi: 10.1287/mnsc.2014.2089

Bodner, R. & Prelec, D. (2002). Self-signaling in a neo-Calvinist model of everyday decision making. Psychology and Economics, Vol I. Brocas and J. Carillo (eds.), Oxford University Press.

Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. Journal of Economic Psychology, 32(1), 73-78. doi: 10.1016/j.joep.2010. 12.001

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547. doi: 10. 1111/jeea.12014

Finch, A. (2011). Experimental Philosophy and the Concept of Moral Responsibility. *Modern Schoolman,* 88 (1/2), 146-160. doi: 10.5840/schoolman2011881/29

García-García, J.A., Reding-Bernal, A., & López-Alvarenga, J.C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Metodología de investigación en educación médica, 2(8)*,217-224.

Jacobsen, C., Fosgaard, T.R., & Pascual-Ezama, D. (2018). Why do We Lie? A Practical Guide to the Dishonesty Literature. *Journal of Economic Surveys*, 32 (2), 357-387. doi: 10.1111/joes.12204

Köbis, N., van der Lingen, S., Cruz, T. D. D., Iragorri-Carter, D., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. (2019). The Look Over Your Shoulder: Unethical Behaviour Decreases in the Physical Presence of Observers. doi: 10.31234/osf.io/gxu96

Kocher, M.G., Schudy, S., & Spantig, L. (2018) I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science,* 64(9), 3995-4008. doi: 10.1287/mnsc.2017.2800

Maggian, V., & Villeval, M.C. (2016). Social preferences and lying aversion in children. *Experimental Economics*, 19, 663-685. doi: 10.1007/s10683-015-9459-7

Markiewicz, L. & Gawryluk, K. (2019). Cheating among children: Temptation, loss framing, and previous cheating. *Journal of Behavioral Decision Making*, 1-15. doi: 10.1002/bdm.2150

Mazar, N., & Ariely, D. (2006). Dishonesty in Everyday Life and Its Policy Implications. *Journal of Public Policy & Marketing* 25(1), 117-126. doi: 10.1509/jppm.25.1.117

Mijović-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 227-240.

Monin, B., & Miller, D.T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81, 33-43. doi: 10.1037/0022-3514.81. 1.33

*Pascual-Ezama, D. Fosgaard, T.R. Cardenas, J.C. Kujal, P. Veszteg, R. Gil-Gómez de Liaño, B. Gunia, B. Weichselbaumer, D. Hilken, K. Antinyan, A. Delnoij, J. Proestakis, A. Tira, M.D. Pratomo, Y. Jaber-López, T. and Brañas-Garza, P.* (2015) Context dependent cheating: Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization,* 116, 379-386.

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(2), 202-217.

Prelec, D., & Bodner, R. (2003). Self-signaling and self-control. In G. Loewenstein, D. Read, & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 277-298). New York, NY, US: Russell Sage Foundation.

Shalvi, S., Dana, J. Handgraaf, M.J.J., & De Dreu, C.K.W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes,* 115(2), 181-190. 10.1016/j. obhdp.2011.02.001

Tembrunsel, A.E., & Messick, D.M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17(2), 223-236. doi: 10.1023/ B: SORE. 0000027411.35832.53

Williams, B. (1981). *Moral Luck*, Cambridge: Cambridge University Press