

ORIGINAL ARTICLE

The effect of observing multiple private information outcomes on the inclination to cheat

Sandro Casal¹ | Antonio Filippin^{2,3}

¹Department of Economics and Management, University of Trento, Trento, Italy

²Department of Economics, Management, and Quantitative Methods, University of Milan, Milan, Italy

³Institute for the Study of Labor (IZA), Bonn, Germany

Correspondence

Sandro Casal.

Email: sandro.casal@unitn.it

Funding information

Università degli Studi di Milano;
Università degli Studi di Trento

Abstract

We investigate how the inclination to cheat changes when agents report the result of multiple realizations of a (private information) stochastic event rather than a single outcome. Multiple realizations render extreme outcomes unlikely, facilitating the identification of opportunistic behaviors and exposing to reputation concerns the individuals who report them. Consequently, multiple realizations lead to a significant reduction of cheating by large amounts. Simultaneously multiple realizations also diminish the intrinsic cost of lying, thereby inducing a widespread inclination to adjust upward the observed outcome in a plausible manner. The overall effect is only a marginal decrease in the degree of cheating.

KEYWORDS

cheating, moral self-licensing, reputation concerns

JEL CLASSIFICATION

C81, C91, D82

1 | INTRODUCTION

Asymmetric information is a crucial issue in economic relationships and has been extensively studied given the harmful consequences that opportunistic behaviors may have on other parties. Contributions have investigated incentives and institutions capable of limiting this problem, but without fully reconciling the conflicting interests involved. For instance, the literature in labor economics has shown that monitoring can be an effective device to reduce informative advantages (e.g., Sappington, 1991; Walker, 2000; Whynes, 1993). However, monitoring suffers some limitations in terms of possible crowding out of intrinsic motivations (e.g., Falk & Kosfeld, 2006; Ichino & Muehlheusser, 2008; Walker, 2000). Moreover, monitoring can be scarcely informative of the workers' behavior when their productivity is jointly determined by both (unobservable) effort and a stochastic component. This is true especially for a single observation of the worker's performance, when the stochastic component may dramatically contribute toward the final outcome.¹

The *employer-employee* is just one among a plethora of possible examples. The possibility to exploit one's informative advantage is in fact a general feature of the principal-agent problem. In all these circumstances truthfully reporting the realization of the stochastic component depends on both monetary and non-monetary incentives of the agent. For this

Managing Editor: Laura Gee

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. Economic Inquiry published by Wiley Periodicals LLC on behalf of Western Economic Association International.

reason, a strand of the experimental literature has recently grown analyzing to what extent agents exploit their private information and what affects their (dis)honest behavior.

As described below in more detail, this literature on lying shows that subjects do not maximize their monetary return when doing so implies an opportunistic behavior. It is worth stressing, however, that while principle-agent relationships usually entail repeated interactions, the literature on lying is mostly based on one-shot games. Some studies deal only with repeated realizations of a random event (Cohn et al., 2014; Lowes et al., 2017; Shalvi, 2012). A few exceptions contain both one-shot and repeated tasks (Abeler et al., 2014; Suri et al., 2011), but their goal is not that of identifying the mere effect of observing multiple realizations. Some indirect evidence suggests that the repetition of a task may indeed induce dishonest behavior, although in experimental settings not meant to investigate this specific issue.²

Identifying how multiple realizations of a stochastic component shape the inclination to exploit one's private information constitutes an original exercise that complements the existing literature and is the main goal of this paper. We envisage two effects. By shaping the ex-ante probability distribution, several independent realizations reveals an opportunistic behavior more easily than a single realization. Borrowing from the literature, we refer to this effect as *reputation concerns*. Back to the monitoring example, observing a combination of outcomes becomes more informative because the plausible range of repeated draws shrinks considerably if the distribution of the stochastic component is known. The employee retains his private information, but shirking in this case increases the non-monetary cost of being perceived as a liar.

Reporting the result of multiple realizations may have a second effect on the inclination to cheat given the same ex-ante probability distribution of the stochastic component. We identify *moral self-licensing* (Monin & Miller, 2001) as the most relevant explanation. According to this theory, being honest most of the times helps the agents to preserve their self-image making an occasional opportunistic behavior morally more acceptable than cheating when reporting a single outcome. Our experimental design is equipped to disentangle these two effects by exogenously manipulating a Cheating Task.

The literature on lying, well summarized by Abeler et al. (2019), finds that subjects do not maximize their monetary return when doing so entails an opportunistic behavior. Analyzing 90 studies, Abeler et al. (2019) report that, on average, subjects tend to misreport their private information and increase their earnings. However, the realized gains are only about 30% of what achievable if they were to cheat to the maximum extent. Abeler et al. (2019) also consider many models—standard (neoclassical) models; lying costs (LC) models; reputation for honesty (RH) models; and Conformity (in lying) models—showing that the data of the meta-analysis are consistent with a combination of LC and RH only. Thus, the experimental evidence can be explained when intrinsic cost of lying and reputation concerns are simultaneously considered. The intrinsic cost of lying can be seen as the individual preference for being honest, since cheating negatively affects one's self-image. This cost is suffered whenever the agent cheats. Reputation concerns may be described as the preference for being seen as honest, because cheating threatens one's social image. Reputation concerns entail a cost that is incurred when the subject knows or suspects that someone else is aware of (or can reasonably infer) his dishonest behavior.

Reporting multiple realizations may trigger opposite effects via self- and social image. As a result, the overall effect of repeating a Cheating Task is not obvious a priori. Surprisingly, there are no studies in the literature that compare in a clean manner whether reporting a stochastic outcome changes when obtained through a single versus multiple realizations. In their meta-analysis Abeler et al. (2019) do not find a significant effect of the repetition of the Cheating Task. However, their findings are based on various studies where other factors vary concurrently with the number of task iterations. In fact, they highlight the absence of an experimental examination specifically targeting the repetition of a Cheating Task.

From a theoretical perspective, three contributions have modeled the non-monetary costs of cheating: Dufwenberg and Dufwenberg (2018); Gneezy et al. (2018); Khalmetski and Sliwka (2019). With specific characteristics in their formalization, all these models share the importance of reputation concerns predicting that individuals cheat less when their claim is more likely to be perceived as dishonest by an observer.³ All these models therefore predict that repeated outcomes should reduce the amount of cheating via reputation concerns.

Differences in the theoretical models are more pronounced as far as the intrinsic cost of cheating is concerned. In Dufwenberg and Dufwenberg (2018) the true realization of the stochastic event does not play any role. Khalmetski and Sliwka (2019) posit that subjects pay a fixed cost of lying, while in Gneezy et al. (2018) the cost also depends on the distance between the report and the true state. Regardless of the specific formalization, the individual cost of lying is procedurally invariant to the number of realizations of the stochastic component in all these models. In other words, whether the outcome to be reported comes from a single or multiple realizations should not affect the intrinsic cost of cheating.⁴

As described in Section 2, we analyze the effect of observing multiple realizations by administering a properly designed Cheating Task in which subjects hold private information about the realization of a stochastic event. The experiment consists of three conditions meant to identify the change in the (intrinsic and social) cost of cheating implied by reporting the result of a single versus multiple realizations. The baseline condition consists in a single realization of a stochastic event drawn from a uniform distribution. The first treatment maintains the single draw but changes the underlying probability distribution in order to make extreme claims implausible. The second treatment manipulates the number of realizations observed, keeping the underlying distribution unchanged. In doing so we pay particular attention in keeping the incentive structure (minimum, expected, and maximum return as well as minimum amount of cheating) constant across conditions.

Our findings, presented in Section 3, show that both the envisaged mechanisms are at work. The first consequence of reporting the result of multiple realizations is that high claims become unlikely and we observe that, coherently, they tend to disappear. In other words, reputation concerns significantly decrease the amount of cheating. There is a second effect, however. The mere fact of observing multiple realizations triggers a more pronounced inclination to cheat by small amounts. Reporting multiple realizations seems to facilitate some cheating occasionally, as if the corresponding intrinsic cost is lowered by behaving honestly most of the times. The two mechanisms operate in opposite directions, with the former being slightly larger in magnitude. As a result, the extent of cheating observed reporting multiple realizations is slightly lower than in the one-shot case, but not significantly so. This result confirms the null difference reported by Abeler et al. (2019). The effect through the erosion of the intrinsic cost of cheating compensates almost entirely the effect of reputation concerns, which at first glance one may expect to dominate.

2 | THE EXPERIMENT

We use a Cheating Task to determine the variable component of the total compensation for a real effort task in which effort is observable. Not implementing the Cheating Task as an end-in-itself assignment has the additional advantage to reduce potential demand effects (Zizzo, 2010). In particular, de Quidt et al. (2018) show that inferring the goal of the study may constitute a demand effect that induces subjects to change their behavior. To further avoid that the Cheating Task alone may reveal the object of the study, we also enrich the protocol including additional tasks meant to elicit potentially relevant explanatory factors of cheating behavior. The first is risk aversion. Given that being considered as a liar is a probabilistic outcome, risk averse subjects may report lower numbers to reduce or eliminate such a risk. Dai et al. (2018) indeed show that risk seeking participant report higher outcomes. We also elicit trust and trustworthiness, that have been shown to positively correlate with honesty by Mann et al. (2016) and, to a lower extent, by Abeler et al. (2014).

In the labor market the total productivity of the worker is typically observable, while its determinants (effort and the stochastic component) cannot be distinguished. Our experimental design implements an isomorphic but cleaner setup, in which effort is observable while the total productivity, which also includes the stochastic component, is reported but not observed. The main reason for this choice is that heterogeneous levels of effort could act as a confounding factor in the cheating decision. For instance, the agents exerting a higher level of effort may feel entitled to exploit their informative advantage to a greater extent. By keeping the effort observable we can check that the design is successful in removing this source of heterogeneity, and use it as an ex-post control if some differences survive.

2.1 | Experimental tasks and design

2.1.1 | Cheating Task

In the Cheating Task subjects are rewarded according to their claim about the realization of a random outcome. The treatments, explained in detail below, consist of reporting this outcome obtained either as the result of a single realization from different distributions, or as the sum of multiple realizations. In all cases the stochastic realization(s) is (are) private information.⁵

Disentangling the effect of the number of realizations observed from that of potential confounding factors requires a careful design. Note that generating multiple outcomes by repeating a classic die roll task à la Fischbacher and Föllmi-Heusi (2013) would not deliver a clean comparison for two reasons. First, avoiding wealth effects requires to pay the

outcome of the one-shot task and the average outcome of the repeated task.⁶ Doing so, however, induces a difference in the minimum amount of cheating allowed. In the repeated task this amount is a fraction ($1/n$, where n is number of repetitions) of its counterpart in the one-shot version. Consequently, a subject whose self-image concerns allow him to cheat only marginally is more likely to misreport the outcome in the repeated than in the one-shot task. Second and foremost, the role played by the number of repetitions would change together with the probability distribution of the random variable.⁷

We implement a task that mimics the rolling of a die once versus 10 times, avoiding such confounding factors. More precisely, the experiment manipulates separately but *ceteris paribus* (i) the number of draws necessary to generate the same random variable, and (ii) the probability distribution of the random variable. The result is an experimental design with three treatments:

1. 1-Uniform (1U). Participants draw only once from a discrete uniform distribution between 10 and 60 and are asked to report the number obtained;
2. 10-Uniform (10U). Participants draw 10 times (with replacement) from a discrete uniform distribution between 1 and 6 and are asked to report the sum of the 10 draws;
3. 1-Normal (1N). Participants draw only once from a distribution generated in the same way as in *10-Uniform* and are asked to report the number obtained.⁸

The number reported represents a stochastic component of the total compensation expressed in experimental currency units (ECU).⁹ Note that the three conditions are isomorphic in terms of monetary incentives. The random variable is identical in terms of support (10–60) and expected outcome (35). Moreover, this setting equalizes the minimum amount of cheating (1 ECU) regardless of the number of draws. The treatments instead manipulate the non-monetary costs of cheating in order to identify the underlying determinants.

The *1-Normal* condition induces a dramatic change in the variance of the outcome with respect to the *1-Uniform*, while keeping constant the fact that subjects observe only a single realization. The comparison of the results in these two conditions allows us to identify the effect of reputation concerns. It is necessary to be careful when relating observed reports that differ from an objective distribution to the notion of lying. For this reason, in the analysis of the data and when comparing results across treatments we adopt a probabilistic approach. Subjects claiming 48 or more in *1-Normal* can be classified as liars with 99% confidence because such outcomes can actually be observed with a cumulative probability of 0.94%. In contrast, no claim could be labeled as mendacious with the same confidence in *1-Uniform* given that even 60 has about a 2% chance of occurring (see Figure 1).¹⁰ The *10-Uniform* treatment manipulates the number of repetitions as compared to *1-Normal*, while keeping constant the underlying distribution of the random variable. Different choices across these two conditions can be attributed to the effect of observing multiple realizations of the stochastic component, where moral self-licensing can operate. The comparison between *1-Uniform* and *10-Uniform* represents the overall effect of observing multiple realizations, including both reputation concerns and moral self-licensing.

2.1.2 | Coin Task

We observe the effort exerted by the subjects using the Coin Task (Gioia, 2016). This task consists of recognizing the nominal value and the issuing country of a sequence of Euro coins. In every round subjects see a table showing 25 coins and reporting their value/country. One of these coins is randomly selected and displayed to the participants, whose task is to identify it correctly (see Figure A1 for an example). The compensation is deterministic and equal to 2 ECU for each coin correctly identified. Participants have 20 min to identify a maximum of 20 coins, with the possibility of making mistakes. Such a long time interval was purposely chosen to facilitate the successful completion of the task, thereby avoiding that wealth effects may act as a confounding factor in the Cheating Task.¹¹

2.1.3 | Bomb risk elicitation task

We elicit subjects' risk preferences using the bomb risk elicitation task (BRET) proposed by Crosetto and Filippin (2013). Participants are presented with a 10×10 square in which each cell represents a box: 99 boxes contain 1

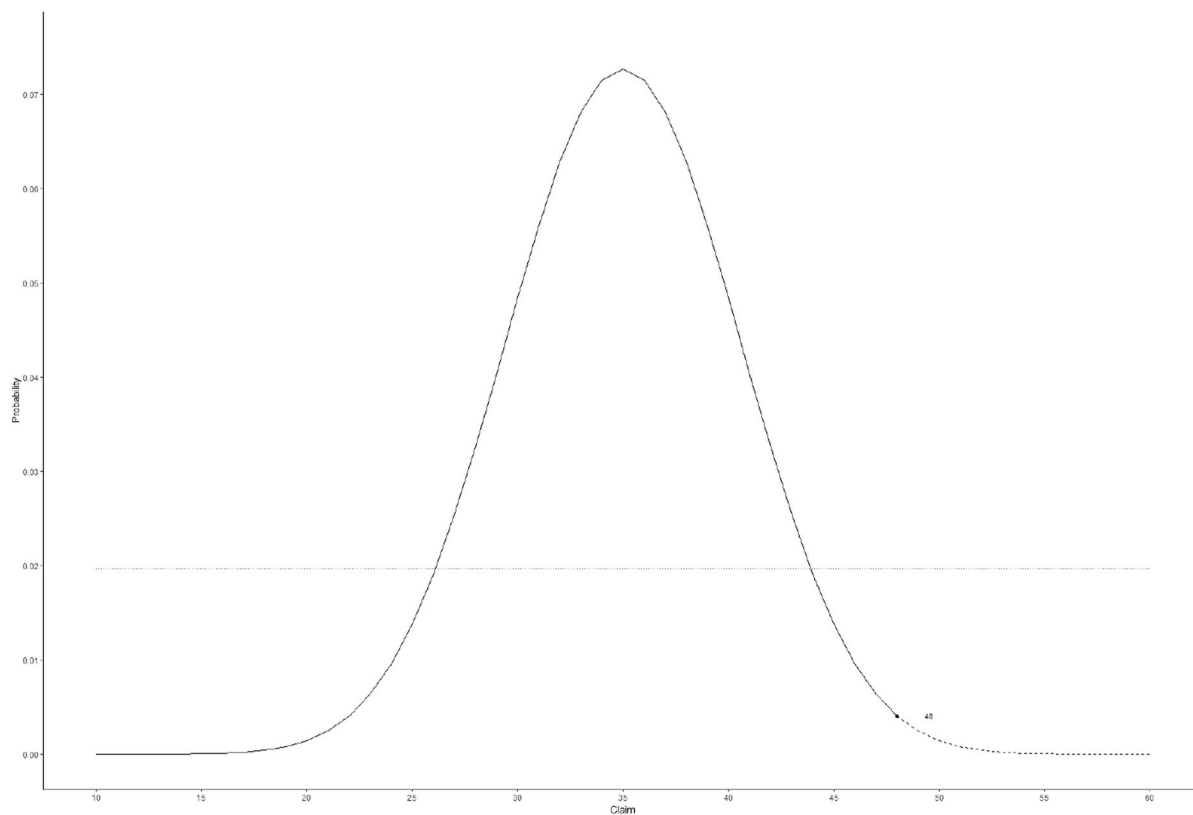


FIGURE 1 Probability distribution of claims. The dotted line displays the uniform distribution in *1-Uniform*, in which each claim has a 1.96% chance of occurring. The solid and dashed line represents the (approximated) normal distribution of outcomes in *10-Uniform* and *1-Normal*. Claims on the dashed part of the line (i.e., starting from 48) can be classified as mendacious with at least 99% confidence.

ECU, while one contains a time bomb. Participants choose how many boxes to collect $k_i \in \{1, 100\}$ knowing that if the bomb is collected earnings will be zero. The position of the bomb, $b_i \in \{1, 100\}$, is randomly determined after the participant's choice. If $k_i \geq b_i$, it means that the subject collected the bomb, which by exploding wipes out the earnings. In contrast, if $k_i < b_i$ the subject receives 1 ECU for every box collected. The number of boxes chosen measures risk preferences: the lower the number, the more risk-averse the subject. $k_i = 50$ represents the risk-neutral choice.

2.1.4 | Mini Trust Game

We propose the Mini Trust Game reported in Figure 2 to measure participants' trust and trustworthiness. Player A chooses between keeping 10 ECU or passing the money to Player B. In the second case the 10 ECU are multiplied by four, and it is then Player B who decides whether to keep the money or split it in a roughly equal way.¹² Player A is defined as trusting when passing the 10 ECU, while Player B is defined as trustworthy when returning 22 ECU back. The game is played only once in strategy method, so that every subject makes a choice as both Player A and Player B. At the end of the experiment participants are randomly matched in pairs and roles are randomly assigned. Individual payoffs are computed by combining participants' decisions in the assigned role.

2.2 | Experimental procedures

The experiment was run in two waves. The first wave took place between January 2018 and February 2019 at the University of Milan (Italy). The treatments in this wave are characterized by a minor difference in the generation of the random variable. In *1-Uniform* and *10-Uniform* the subjects physically draw the numbers from a bag containing tokens. In *1-Normal* a physical replication of the underlying distribution is not feasible because it would require several thousands of tokens. Hence, subjects draw the number by accessing simultaneously an external website with their

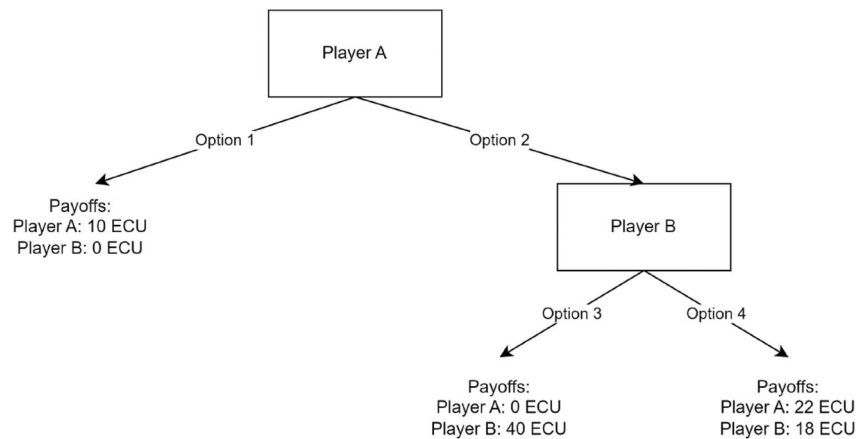


FIGURE 2 Mini Trust Game.

smartphones. It is important to notice that, in both procedures, the generation of the random variable is disconnected from the computer where the subjects report the number observed. While there is no meaningful way in which observed and reported numbers can be connected ex-post, in *1-Normal* subjects may wrongly perceive a relatively stronger degree of observability of their choices.¹³

In order to rule out any possible effect of such a difference, a second wave was programmed, but could take place only after the pandemic, between April 2022 and May 2023, at the University of Trento (Italy). In the second wave subjects draw the random numbers accessing privately different third-party websites in all treatments.¹⁴ In Section 3 we show that the results by treatment do not significantly differ across waves. In Wave 1 subjects were recruited with ORSEE (Greiner, 2015) and the experiment was programmed and conducted using z-Tree (Fischbacher, 2007). In Wave 2 subjects were recruited with an internal recruitment-system, and the experiment was programmed and conducted using o-Tree (Chen et al., 2016). In both waves the participants were university students.

Upon their arrival, participants were randomly assigned to one of the laboratory seats, all separated by partitions. Participants were informed that the experiment was composed by three independent phases, with the respective instructions (see Appendix A) read aloud at the beginning of each phase:

1. BRET;
2. Mini Trust Game;
3. Real Effort Task (Coin Task + Cheating Task).

At the beginning of Phase 3, subjects were informed that the earnings of the real effort task phase had two components: a deterministic part (the score in the Coin Task) and a stochastic part (unrelated to the number of coins correctly identified).

The three phases were preceded by a short questionnaire in which, besides some demographics, we collected information on subjects' familiarity with joint probabilities (details in Appendix B). For reputation concerns to matter, in fact, a necessary condition is that subjects understand that drawing a high number in *1-Normal* and *10-Uniform* is less likely than in *1-Uniform*.

The experimental protocol is identical in all sessions as far as the questionnaire, the BRET, the Mini Trust Game, and the Coin Task are concerned. The only difference is the between-subjects manipulation of the Cheating Task (*1-Uniform*, *10-Uniform*, *1-Normal*). In *10-Uniform* subjects have to report the sum of 10 draws. In order to avoid a potential confounding factor due to a higher cognitive load, they are told that they can write down the outcomes using their smartphones or a piece of paper. To prevent perceived differences in the underlying distributions, subjects in *1-Normal* are reminded that the number they are going to see is the sum obtained by drawing 10 times a number from 1 to 6 with replacement. In this way subjects in *1-Normal* get the same information about how the distribution is generated as those in *10-Uniform*.

The experiment ended with the resolution of uncertainty of the first two phases and the payment. In addition to a show-up fee, participants received the payoff of each phase. The pay-all-tasks protocol is the most appropriate in this

experiment because the compensation for the overall productivity needs to encompass both the effort exerted and the stochastic component.¹⁵ For the sake of salience of the non-monetary incentives, payments are made privately but directly by the experimenters. On average, sessions lasted 50 min and the average earnings amounted to €8.60.

3 | RESULTS

A total of 668 subjects took part in the experiment, 360 in the first wave and 308 in the second. Table 1 shows that the claims in the Cheating Task do not significantly differ across waves, in particular where the randomization device was not the same (*1-Uniform* and *10-Uniform*). In what follows, we therefore analyze the data merging the two waves.

Table 2 presents the descriptive statistics of the sample across treatments, with participants divided as follows: 286 in the *1-Uniform*, 192 in the *10-Uniform*, and 190 in the *1-Normal* treatment. The larger sample in *1-Uniform* is due to the higher variance of the outcomes. The groups are balanced from a gender perspective. As far as the behavioral traits are concerned, trust and trustworthiness are indistinguishable. The randomization across treatments deliver results that are less clear as far as risk preferences are concerned. Participants in *1-Normal* turn out to be more risk-averse, although the differences are not significant. In any case, as discussed in detail in Section 3.3, this variable does not play a significant role in explaining the inclination to cheat and therefore it does not constitute a confounding factor. All participants achieve the maximum score in the Coin Task, ensuring that claims in the Cheating Task are not influenced by different monetary rewards.

The average claim in the Cheating Task, our main variable of interest, is 38.96. Hence, subjects lie on average since claims are significantly higher than the expected value of 35 (Mann–Whitney test, $p < .001$). However, as commonly found in the literature, subjects exploit their private information only to a low extent. The claim in excess, computed as the difference between participants' reported outcomes and the theoretical expectation, is only 3.96 ECU on average, that is, about 16% of the maximum amount (25 ECU). The non-monetary cost of cheating confirms to be an extremely important bulwark against opportunistic behavior.

TABLE 1 Claims in the Cheating Task across waves.

Treatment	Wave 1		Wave 2		p-value (MW)
	Average	N	Average	N	
1-Uniform	39.76	182	41.53	104	.406
10-Uniform	39.27	88	39.43	104	.761
1-Normal	35.98	90	36.72	100	.432
Total	38.70	360	39.26	308	.669

Abbreviation: MW, Mann–Whitney U test.

TABLE 2 Balancedness across treatments.

	Treatment			p-value (test)		
	1-U	10-U	1-N	1-U versus 10-U	1-U versus 1-N	10-U versus 1-N
No. of subjects	286	192	190			
% of females	55.59	57.81	62.10	1.000 (FE)	0.552 (FE)	1.000 (FE)
% Trusting	64.33	60.94	63.15	1.000 (FE)	1.000 (FE)	1.000 (FE)
% Trustworthy	52.10	49.48	54.21	1.000 (FE)	1.000 (FE)	1.000 (FE)
BRET	41.47	42.00	38.41	1.000 (MW)	0.090 (MW)	0.147 (MW)
# of coins	20	20	20	-	-	-
Cheating Task	40.41	39.36	36.37	0.123 (MW)	<0.001 (MW)	<0.001 (MW)

Note: The p -values are adjusted using the Bonferroni correction to account for the fact that our design involves three pairwise comparison across treatments. Abbreviations: 1-N, 1-Normal; 1-U, 1-Uniform; 10-U, 10-Uniform; FE, Fisher's Exact test; MW, Mann–Whitney U test.

Table 2 shows that the average claim in *1-Uniform* is not statistically different from that in *10-Uniform*. This result corroborates the evidence provided by Abeler et al. (2019) about the null effect of repeating the die roll task. What they find comparing different papers holds true when eliminating potential confounding factors and within a homogeneous subject pool. This finding is counterintuitive at first glance because reporting the sum of multiple random draws has only a marginal impact on the average degree of cheating despite the much stronger reputation concerns characterizing the *10-Uniform* treatment.

The value added of our experiment is the possibility to explain what drives this null result. The *1-Normal* treatment allows us to establish that the similar outcome in the other conditions is due to the composition of two significant effects, which operate in opposite directions: reputation concerns and moral self-licensing driven by observing multiple realizations. We can see in Table 2 that claims in *1-Normal* are indeed significantly lower than in *1-Uniform* and *10-Uniform*, and not far from the theoretical prediction without cheating.

3.1 | Reputation concerns

We focus first on reputation concerns, which in our setting can be identified by comparing the *1-Uniform* and the *1-Normal* treatments. In both conditions subjects perform a single random draw. However, by reducing the range of outcomes that can be reasonably expected, the *1-Normal* condition makes an opportunistic behavior much easier to detect at the individual level. The disutility suffered when others (the experimenters in this case) may believe that their behavior is dishonest should affect subjects' claims significantly more in *1-Normal* than in *1-Uniform*. Since all the other determinants (monetary incentives, intrinsic cost of lying, number of realizations observed) are constant, different claims in these two treatments identify the intensity of reputation concerns. Claims are indeed significantly higher in *1-Uniform* than in *1-Normal* (40.41 vs. 36.37). We therefore find that reputation concerns matter.

The average claim is a useful statistic, but it overlooks a lot of relevant information. For instance, is the observed difference driven by a slight but widespread change in the behavior, or does it follow from an effect concentrated on the highest claims? A first answer to this question is provided by Figure 3, which describes the observed distribution of claims compared with their theoretical counterparts (dotted lines). The observed claims in *1-Normal* (Figure 3a) follow rather closely the underlying theoretical distribution. As regards *1-Uniform* (Figure 3b), the distribution is instead markedly skewed to the left. The weak reputation concerns facilitate an opportunistic behavior that is more evident looking at the highest claims. At the same time we still observe a substantial fraction of low (and presumably truthful) claims.

Before analyzing the effect of reputation concerns in more detail we need to clarify what we deem as an opportunistic behavior at the individual level. Imagine a subject observing a realization of the random variable. While this subject suffers the intrinsic cost of cheating whenever reporting more than the number observed, he suffers reputation concerns only if (he suspects that) others believe that the claim is dishonest. Hence, it is crucial to define which claims are “sufficiently high” to be (or at least feared to be) meaningfully labeled as mendacious. Given that any rule to classify

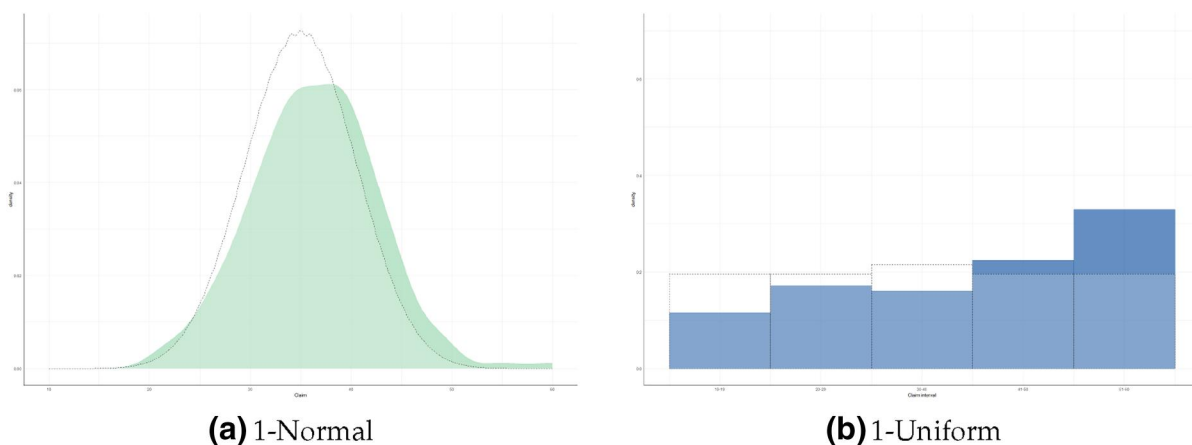


FIGURE 3 Reputation concerns: density of claims by treatment. For graphical convenience, we report the kernel density of observed claims in *1-Normal* (Panel a) and histograms (collapsed in 5 bins) for *1-Uniform* (Panel b). Dotted lines represent the expected frequencies.

a claim as dishonest intrinsically contains a certain degree of arbitrariness, we decide to rely on the conventional 95% confidence level. This approach identifies the range 45–58 as the one in which claims are unlikely to occur only in the *1-Normal* condition. In fact, claims up to 44 leave more than a 5% probability that a greater or equal outcome is actually observed in *1-Normal*. Consequently, the probability of a type I error is too high to reject the null assumption of honest reporting. Conversely, 59 and 60 constitute unlikely claims in *1-Uniform*, too. Hence, the 45–58 range is where the effect of reputation concerns should be reasonably observed.

The next step is to estimate the fraction of cheaters along this distribution of claims. In doing so we have to take into account the different objective probabilities of each outcome across the two treatments (see Figure 1). A comparable metric is provided by the excess percentage of subjects reporting more than a number divided by the percentage that should not have observed it, given the objective probability (Abeler et al., 2019). For instance, in *1-Uniform* 181 subjects out of 286 (63.3%) claim strictly above 35. However, only about 49% should have actually observed a value above 35. Hence, we observe 14.3% of claims higher than 35 coming from the 51% of the population that should have observed an outcome up to 35. Therefore, $14.3\%/51\% = 28\%$ is the estimated percentage of cheaters. Garbarino et al. (2018) improves this statistic developing a method that weights all the possible realizations of the random variables according to their objective probability of occurring, rather than simply imputing the expected percentage. In the example above, the method iterates the computation of the percentage of cheaters considering the objective probability that any number of subjects, from 0 to 181, actually observed an outcome higher than 35.¹⁶ Another advantage of this procedure is that it provides an estimate of the probability distribution of the fraction of cheaters rather than just a point estimate.

Table 3 reports for every threshold between 45 and 58:

1. The estimated fraction of cheaters by treatment, calculated using the software developed by Garbarino et al. (2018)¹⁷;
2. The *p*-value of a Chi-square test for the difference in the proportions, adjusted using the Bonferroni correction to account for the fact that our design involves three pairwise comparison across treatments.

In *1-Uniform* each claim is equally likely. Interestingly, claims in the range 45–58 are associated to a fraction of cheater that decreases from 20.62% to 6.82% as we approach the upper bound of the range where reputation concerns should matter. In *1-Normal* claims over the whole interval 45–58 signal an opportunistic behavior. Indeed, we observe fraction of cheaters that is much lower than in *1-Uniform* and becomes fairly close to zero starting from 49. As a result, the estimated fraction of cheaters is significantly higher in *1-Uniform* than in *1-Normal* for every threshold in the range 45–58. Summarizing, we find compelling evidence that the likelihood of being detected as a liar triggers reputation concerns.

Result 1. *When indicative of opportunistic behavior, the occurrence of high claims is significantly lower.*

3.2 | Multiple observations

The main research question of the paper is to investigate whether the mere observation of multiple stochastic realizations plays an independent role, establishing whether also this channel affects the individuals' claims. This effect

TABLE 3 Reputation concerns.

	Threshold													
	45	46	47	48	49	50	51	52	53	54	55	56	57	58
% of cheaters														
1-Normal	2.83	2.68	2.76	2.63	0.78	0.90	0.98	1.02	1.03	1.04	0.52	0.52	0.52	0.52
1-Uniform	20.62	20.85	21.06	20.81	19.67	16.42	15.44	14.10	12.81	12.77	11.58	9.67	8.59	6.82
<i>p</i> -value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.006

Note: Each column presents the estimated fraction of cheaters, estimated following Garbarino et al. (2018), for that specific threshold. The *p*-value (adjusted with the Bonferroni correction) refers to a Chi-square test for the difference in the proportions across treatments.

can be identified in our experimental setting by comparing *10-Uniform* and *1-Normal*. In these two treatments subjects report only one number, whose objective distribution is identical. The distribution is obtained in both cases drawing 10 times randomly—with replacement—a number from 1 to 6. Extreme claims have the same (low) probability of occurring and reputation concerns are therefore constant. Any difference in the choices must depend on the way the final outcome is generated. In *10-Uniform* the subjects perform the 10 random draws, while in *1-Normal* they draw only once from a distribution generated with the same procedure. As already shown in Table 2, claims are indeed significantly higher in *10-Uniform* than in *1-Normal* (39.36 vs. 36.37). We therefore find that observing multiple realizations increases the inclination to cheat. This effect has the opposite sign and is roughly comparable in size to the role played by reputation concerns.

Also in this case we try to shed some light on the underlying mechanism going beyond the analysis of average claims. Figure 4 reports the observed claims compared with the same theoretical distribution (dotted lines) in the two conditions. The vertical distance between the solid and dotted lines provides a prima facie indication of the amount of cheating along the distribution. It is immediately evident that the distribution of claims in *10-Uniform* (Figure 4b) appears shifted to the right as compared to *1-Normal* (Figure 4a). The interpretation is that in *10-Uniform* reputation concerns are still somehow effective in hindering very high claims. However, observing multiple outcomes seems to induce a widespread misreporting toward relatively higher values.

This speculation is confirmed by looking at the proportion of participants claiming values strictly higher than the average theoretical claim (35). Such a fraction is 54.2% in *1-Normal* and 67.2% in *10-Uniform* (Fisher exact test, $p = .012$). We implement again the Garbarino et al. (2018) method to estimate the fraction of cheaters along the whole distribution. Table 4 shows that multiple observations of the stochastic event induce a widespread opportunistic behavior, with the fraction of cheaters that results significantly higher in *10-Uniform* across the board.¹⁸

Two effects are worth being noted. First and foremost, the percentage of cheaters is significantly higher in *10-Uniform* along the whole distribution, starting from very low thresholds. Second, the fraction of cheaters in *10-Uniform* decreases progressively, but without converging to zero. In other words, a few subjects seems to be insensitive to reputation concerns. To test whether these subjects rather than insensitive are simply unaware of reputation concerns we use the information collected in the pre-experimental questionnaire. We ask the probability to obtain a sum equal to 2 and to 7 rolling two dice (more details in Appendix B) and classify subjects as *unfamiliar* with the logic of joint probabilities if they do not recognize at least qualitatively that obtaining a 7 is more likely than a 2. Regardless of their answer, participants receive the solution of the two questions and a brief explanation of why it is more likely to obtain 7 than 2. Possibly for this reason we find that claims do not differ according to the ex-ante (un)familiarity, as shown by Table 5. Furthermore, among subjects claiming at least 44, that is, about the interval where reputation concerns significantly matter, the fraction of *familiar* is even higher in *10-Uniform* (69.6%) than in *1-Normal* (52.2%) though not significantly so (Fisher exact test, p -value = .189).

Repeated observations seem therefore to erode the intrinsic moral cost of the subjects, fostering partial lying. On average, subjects tend to readjust their outcome slightly upward, while still preserving their social image.

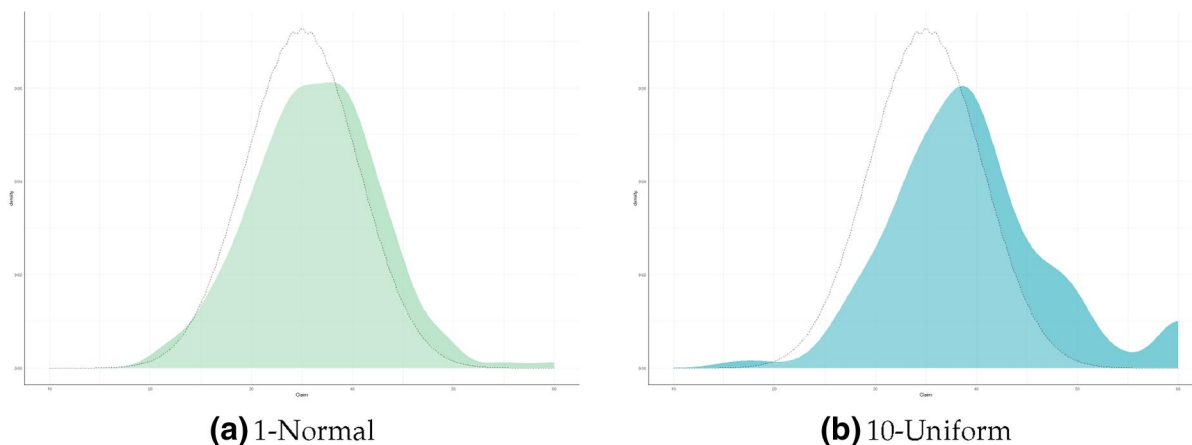


FIGURE 4 Multiple observations: density of claims by treatment. Kernel density of observed claims in both conditions: *1-Normal* (Panel a) and *10-Uniform* (Panel b). Dotted lines represent the expected frequency.

TABLE 4 Observing repeated realizations.

	Threshold															
	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
% of cheaters																
1-Normal	18.64	17.73	20.01	19.08	17.92	14.45	16.62	17.27	15.08	12.50	10.06	7.66	6.94	6.66	3.62	
10-Uniform	45.51	47.34	46.35	47.77	42.35	38.54	40.67	38.98	32.20	30.46	24.59	23.35	20.40	19.24	17.60	
<i>p</i> -value	.063	.009	.012	<.001	.003	<.001	<.001	<.001	.003	<.001	.003	<.001	<.001	.003	<.001	
	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	
% of cheaters																
1-Normal	2.83	2.68	2.76	2.63	0.78	0.90	0.98	1.02	1.03	1.04	0.52	0.52	0.52	0.52	0.52	
10-Uniform	17.16	15.32	13.77	12.02	10.15	7.15	6.70	6.21	6.23	5.72	5.20	5.20	5.20	5.20	5.20	
<i>p</i> -value	<.001	<.001	<.001	.003	<.001	.015	.024	.045	.045	.075	.045	.045	.045	.045	.045	

Note: Each column presents the estimated fraction of cheaters, estimated following Garbarino et al. (2018), for that specific threshold. The *p*-value (adjusted with the Bonferroni correction) refers to a Chi-square test for the difference in the proportions across treatments.

TABLE 5 Claims according to familiarity with joint probabilities.

	Claim		<i>p</i> -value (MW)
	Familiar	Unfamiliar	
1-Uniform	40.60	40.17	0.756
10-Uniform	39.50	39.18	0.7006
1-Normal	36.21	36.64	0.6985

Abbreviation: MW, Mann-Whitney *U* test.

Result 2. Observing multiple realizations of a random variable increases the inclination to cheat, relaxing the intrinsic cost of lying.

We found three explanations in the literature that are broadly consistent with the interpretation that multiple observations reduce the intrinsic cost of cheating. The first is that subjects tend to indulge in opportunistic behavior as long as their lies are justifiable both in the eyes of others and to themselves (Shalvi et al., 2011). This self-justification is intended as *observed desired counterfactuals* and has been shown by Shalvi et al. (2011) to operate in repeated tasks. For instance, subjects might feel justified when changing some unlucky draws with higher values that they have actually observed. Although appealing, we believe that this explanation does not fit well to our setting for two reasons. First, in *10-Uniform* each draw cannot be meant as a counterfactual because all the realizations count toward the final payoff. Second, it is difficult to argue that lies are more justifiable to others in *10-Uniform* than in *1-Normal* because the underlying distribution is the same.

The second explanation refers to *ethical blind spots*, which could rationalize a more pronounced cheating in *10-Uniform* with an attentional bias (Pittarello et al., 2015). According to this justification subjects' attention may focus on pleasant information (high numbers) during the repetition of the task, while overlooking bad outcomes (low numbers). Although this mechanism can potentially operate only when multiple outcomes are observed, that is, in *10-Uniform*, we do not find it fully convincing either. An upward adjustment of the sum requires an active effort to substitute low numbers with higher ones. An attentional bias that simply disregards unpleasant information would not be enough.

The third explanation refers to the so-called *moral self-licensing* effect (Monin & Miller, 2001) and is in our opinion the most convincing one. This explanation posits that subjects are more likely to cheat when their misconduct can be compensated by behaving in a moral way in other circumstances, something that in our framework can only occur when observing multiple random draws.¹⁹ A possible pattern in *10-Uniform* is that after taking truthful note of some outcomes, subjects may feel entitled to adjust upwards the result in a subsequent round. Mendacious behavior may then induce them to return to an honest attitude in order to restore their self-image (moral cleansing effect, Blanken et al., 2014).²⁰

The exact sequence is not important. While *moral self-licensing* is the explanation that better rationalizes our findings, our experiment is not equipped to delve into the specific dynamics that the *10-Uniform* treatment may entail.²¹ If subjects are honest most of the times while occasionally indulging in opportunistic behavior, the implication is in any case that the final claim should exceed the actual sum only by few units. In fact, in a single round the upward adjustment is limited by the fact that the maximum number is 6. The implications of this mechanism are in our opinion consistent with the widespread increase of partial lying that we observe.

3.3 | The overall effect

The comparison between *1-Uniform* and *10-Uniform* replicates the one-shot versus repeated die roll paradigm with identical incentives. Our results in Table 2 show that the amount of cheating is slightly higher (1 ECU) in *1-Uniform*, though not significantly so (p -value = .123, adjusted for multiple treatments). This evidence confirms by and large the null difference reported by Abeler et al. (2019) when comparing heterogeneous studies.

Our results in the previous sections allows us to rationalize this result as the composition of two opposite effects of the repetition of the task. On the one hand, we do observe a sizeable and significant effect of the social cost of cheating (via reputation concerns) that prevents very high claims that would otherwise be observed. On the other hand, multiple observations decrease (likely via moral self-licensing) the intrinsic cost of cheating, leading to a more pervasive inclination to readjust the outcome slightly upward. The magnitude of the first effect turns out to be marginally larger than the second. As a result, the amount of cheating observed reporting repeated outcomes is only slightly lower than that observed in the one-shot case. Essentially, observing multiple random events slightly reduces the amount of cheating as a result of: (i) preventing cheating by very large amounts thanks to reputation concerns, and (ii) increasing partial lying because of a lower intrinsic cost of cheating.

We now investigate the determinants of subject's claim in a multivariate analysis. Table 6 presents the results of different linear regression models in which we use subjects' observable characteristics and their behavioral traits as explanatory variables. The treatment effects are captured by the *1-Uniform* and *10-Uniform* dummies. Their coefficients identify the difference in claims as compared to omitted condition *1-Normal*. Column 1 trivially confirms that the effect both of reputation concerns and of observing multiple realizations is strongly significant. Column 2 adds as explanatory variables the elicited behavioral traits. In more detail, *Trust* is equal to 1 when the participant passes the 10 ECU as Player A in the Mini Trust Game. *Trustworthy* is equal to 1 when the participant returns 22 ECU as Player B. *Risk Tolerance* is the choice in the BRET. Trusting, trustworthy, and risk averse subjects tend to cheat less. While the coefficients of these behavioral trait take the expected sign, only *Trustworthy* turns out to be marginally significant. The treatment effects are robust to the inclusion of these controls. In particular, Column 2 shows that the treatment effects are genuine and that risk aversion does not act as a confound. Columns 3 and 4 replicate the analysis while also including a gender dummy. The results above hold virtually unchanged in terms of size and significance while females claim significantly less than males on average, in line with what was previously found in the literature (Abeler et al., 2019).

4 | CONCLUSION

This paper investigates experimentally how observing a single versus multiple realizations of a random variable affects the inclination to cheat when agents hold private information. The starting point is the result in the meta-analysis by Abeler et al. (2019), who find that the amount of cheating does not significantly differ when repeating a classic die roll paradigm. This interesting but counterintuitive result comes from the aggregation of heterogeneous studies, thereby leaving unanswered the analysis of the causal relationships underneath.

Two channels operate at the same time when simply repeating a classic die roll paradigm, on top of possible confounding factor in terms of incentive structure. The first is more intuitive and is represented by the change in the underlying distribution of the random variable. If the distribution of the stochastic component is common knowledge, observing multiple realizations makes extreme outcomes less likely, thereby signaling to the principal that the claim is not truthful. Agents who have reputation concerns should therefore avoid such claims. The second is more subtle and has to do with the mere observation of repeated realizations, given the same underlying distribution. This channel has nothing to do with the principal, and must therefore be related to the intrinsic cost of cheating.

TABLE 6 Linear regression models.

	Dependent variable			
	Claim in the Cheating Task			
	(1)	(2)	(3)	(4)
1-Uniform	4.035*** (1.047)	3.882*** (1.045)	3.873*** (1.043)	3.749*** (1.042)
10-Uniform	2.986*** (1.144)	2.736** (1.144)	2.879** (1.139)	2.659** (1.139)
Trusting		-0.905 (0.971)		-0.923 (0.967)
Trustworthy		-1.664* (0.940)		-1.456 (0.940)
Risk		0.042 (0.027)		0.039 (0.027)
Female			-2.492*** (0.874)	-2.262*** (0.875)
Constant	36.374*** (0.811)	36.230*** (1.464)	37.921*** (0.972)	37.656*** (1.559)
Observations	668	668	668	668
R ²	.022	.036	.034	.045
Adjusted R ²	.019	.028	.030	.037

Note: 1-Uniform and 10-Uniform are treatment dummies. Trusting is a dummy equal to 1 when the subject passes the 10 ECU in the Mini Trust Game. Trustworthy is a dummy equal to 1 when the subject returns the 22 ECU in the Mini Trust Game. Risk is the number of boxes collected in the BRET. Female is a gender dummy.

Abbreviations: BRET, bomb risk elicitation task; ECU, experimental currency units.

* $p < .1$, ** $p < .05$, *** $p < .01$.

A parallel situation can be observed in the labor market. When workers' productivity is jointly determined by effort and by a stochastic component, observing a single outcome can be uninformative. As long as the stochastic component may account for the final outcome over and above the worker's effort there is much room for shirking. The monetary incentives of a misconduct must be weighted only against the self-image cost of dishonest behavior. Multiple outcomes are instead more informative, since a misconduct also implies the additional reputation cost of being detected as a liar. Concurrently, reporting multiple outcomes may also affect the intrinsic cost of cheating.

Toward this goal we set up two treatments identifying precisely these two different channels through which multiple outcomes may play a role, as compared to a baseline condition in which we administer a classic (and one-shot) Cheating Task. We do so paying great attention to maintaining the structure of monetary incentives (including expected outcomes and their support, as well as the minimum amount of cheating allowed) identical across all the experimental conditions.

We find that reporting the sum of multiple realizations rather than a single realization induces two opposite effects. On the one hand, it induces reputation concerns that dramatically reduce the occurrence of severe misreporting. Knowing that extreme outcomes are unlikely, subjects tend to avoid the social-image cost of making implausible claims.

On the other hand, reporting the sum of multiple realizations seems to decrease the intrinsic cost of cheating, leading to a more pervasive inclination to adjust slightly upward the true realization. Moral self-licensing is the explanation that better fits our results. Behaving in a honest way most of the time helps to preserve one's self-image even when behaving in an opportunistic way occasionally. The second effect is small in magnitude but widespread. At the aggregate level it compensates to almost entirely the effect of reputation concerns.

As a result, the overall effect is null, confirming the result reported by Abeler et al. (2019) about the repetition of the task. The effect through the erosion of intrinsic costs compensates almost entirely the effect of reputation concerns, which at first glance one may expect to dominate. The parallel example in the labor market is that having workers reporting their productivity over a sufficiently long period of time does not have an obvious impact. In fact, the positive effect on productivity triggered by reputation concerns should be carefully weighted against the opposite force represented by a stronger inclination to make slightly higher but plausible claims that do not signal an opportunistic behavior.

ACKNOWLEDGMENTS

We are grateful to the University of Milan and to the University of Trento for financial and logistic support. We would like to thank the editors, the anonymous referees, Maria Bigoni, Martin Dufwenberg, Caterina Giannetti, Matteo Ploner, as well as the participants to the first and second BEEN workshop (Bologna and Rome), the ninth SEET Conference (Lecce), the 2018 ESA World Meeting (Berlin), the 13th Nordic Conference on behavioral and Experimental Economics (Odense), and the Euregio Economics Workshop 2019 (Trento) for their useful suggestions. All remaining errors are ours.

DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are openly available in OPENICPSR at <https://doi.org/10.3886/E194601V3>, Casal and Filippin (2023).

ENDNOTES

- ¹ Feedback systems such as the (un)smiley face buttons in airports, stations, malls, etc. are examples of imperfect monitoring of productivity.
- ² For instance, Fischbacher and Föllmi-Heusi (2013) find that subjects participating for the second time in the same Cheating Task claim higher payoffs than subjects without previous experience. In situations where subjects are required to report only the first among multiple realizations Shalvi et al. (2011) and Gino and Ariely (2012) find that favorable results in non-relevant rounds make subjects feel justified for acting in a more opportunistic manner. Braut and Piovesan (2020) show that subjects claim more when reporting the last out of five numbers observed than when reporting their sum.
- ³ Gneezy et al. (2018) corroborate experimentally the role of reputation concerns by finding that partial lying occurs only when the subjects hold private information. By not claiming the highest possible payoff, subjects try to preserve their social image leaving their dishonest behavior undetected. In contrast, when the behavior is directly observed by the experimenter the fewer subjects who lie (and therefore who do not care about their social image) tend to do so to the maximum extent. Similarly, Abeler et al. (2019) find that direct observability strongly reduces cheating.
- ⁴ Only in case the cost refers to the observation of each single realization of the stochastic component rather than to the report of the cumulative result, the intrinsic cost of cheating would be higher under multiple realizations because the fixed cost would be paid repeatedly.
- ⁵ For the sake of comparability across treatments subjects report only one outcome (the sum) even when observing multiple realizations. Gerales et al. (2019) show that manipulating the number of outcomes reported given the same number of realizations observed does not alter the choices significantly.
- ⁶ The alternative of paying one outcome at random in the repeated task is not viable for two reasons. On the one hand, it would imply the comparison of a sure amount in the one-shot task with a lottery in the repeated version, with potential consequences driven by risk preferences. On the other hand, the pay-one-at-random protocol is suitable for independent decisions, while reporting repeated outcomes entails intertwined choices.
- ⁷ These caveats clarify why the comparison of one-shot versus repeated tasks presented in the meta-study of Abeler et al. (2019) cannot deliver conclusive evidence.
- ⁸ The distribution of the sum of 10 random draws from 1 to 6 approximates a normal distribution, and therefore we use the term *1-Normal* for the sake of simplicity.
- ⁹ The exchange rate is 20 ECU = 1 Euro.
- ¹⁰ Using a lower confidence level (95%) the thresholds would be 45 in *1-Normal* and 59 in *1-Uniform*. Regardless of the confidence level chosen, we cannot expect the subjects to know exactly the interval of claims that significantly signal a dishonest behavior, particularly in *1-Normal*. Hence, in Section 3 we analyze the lying behavior using the full range of claims for which *1-Normal* entails significant reputation costs (at 95% confidence) while *1-Uniform* does not.
- ¹¹ As shown in Section 3 all subjects indeed successfully completed the task.

- ¹² As noted by Ermisch and Gambetta (2006), this structure of the game without equal payoffs after Option 4 prevents choices from being driven by reasons of fairness. See Ermisch et al. (2009) for a detailed description of why this structure of the game is particularly appropriate for capturing trust and trustworthiness.
- ¹³ Some evidence suggests that reputation concerns may also be triggered by subtle virtual cues of observability (Mol et al., 2020) as well as by different levels of perceived observability (Lilleholt et al., 2020).
- ¹⁴ The website used (last accessed: December 12, 2023) are: <https://www.calculator.net/random-number-generator.html> for treatment *1-Uniform*, <https://www.calculator.net/dice-roller.html> for *10-Uniform*, and <https://pinetools.com/gaussian-random-number-generator> for *1-Normal*.
- ¹⁵ Wealth effects are not an issue in our setting. On the one hand, earnings in the Coin Task are equal for all the participants. On the other hand, the resolution of uncertainty in the first two phases occurs only at the end of the whole experiment and subjects cannot secure any positive amount with their choices.
- ¹⁶ The upper bound is 181 because of the assumption that subjects do not lie downward.
- ¹⁷ Software available at <http://lyingcalculator.gate.cnrs.fr/>. Last access: December 12, 2023.
- ¹⁸ Thresholds below 30 are not reported because the number of potential cheaters is too low to make any meaningful inference. Sixty is also excluded because it is not possible to report a higher payoff.
- ¹⁹ See Blanken et al. (2015) and Effron and Conway (2015) for comprehensive reviews of the argument or Ploner and Regner (2013) and Clot et al. (2014) for examples of moral self-licensing in cheating games. Although evidence of such an effect comes primarily from the experimental literature, moral self-licensing is a robust phenomenon that also occurs outside the laboratory (Hofmann et al., 2014).
- ²⁰ Barron (2019) finds that subjects even lie downward in low-stakes situations in order to signal honesty and to mitigate the repercussions of upward lying in high-stakes contexts.
- ²¹ Some contributions in the literature posit the so-called “slippery slope,” that is, that cheating should increase over time (Engelmann & Fehr, 2016; Garrett et al., 2016). Braut and Piovesan (2020) do not find evidence supporting the slippery slope hypothesis.

REFERENCES

- Abeler, J., Becker, A. & Falk, A. (2014) Representative evidence on lying costs. *Journal of Public Economics*, 113, 96–104. Available from: <https://doi.org/10.1016/j.jpubeco.2014.01.005>
- Abeler, J., Nosenzo, D. & Raymond, C. (2019) Preferences for truth-telling. *Econometrica*, 87(4), 1115–1153. Available from: <https://doi.org/10.3982/ecta14673>
- Barron, K. (2019) *Lying to appear honest*. Discussion Paper 2019-307. WZB.
- Blanken, I., van de Ven, N. & Zeelenberg, M. (2015) A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4), 540–558. Available from: <https://doi.org/10.1177/0146167215572134>
- Blanken, I., van de Ven, N., Zeelenberg, M. & Meijers, M.H. (2014) Three attempts to replicate the moral licensing effect. *Social Psychology*, 45(3), 232–238. Available from: <https://doi.org/10.1027/1864-9335/a000189>
- Braut, B. & Piovesan, M. (2020) *How does dishonesty split? Testing singular vs. multiple incentives*. Mimeo.
- Casal, S. & Filippin, A. (2023) *ECIN replication package for “The effect of observing multiple private information outcomes on the inclination to cheat”*. Ann Arbor: Inter-University Consortium for Political and Social Research [distributor]. Available from: <https://doi.org/10.3886/E194601V3>
- Chen, D.L., Schonger, M. & Wickens, C. (2016) oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97. Available from: <https://doi.org/10.1016/j.jbef.2015.12.001>
- Clot, S., Grolleau, G. & Ibanez, L. (2014) Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters*, 123(2), 191–194. Available from: <https://doi.org/10.1016/j.econlet.2014.01.039>
- Cohn, A., Fehr, E. & Maréchal, M.A. (2014) Business culture and dishonesty in the banking industry. *Nature*, 516(7529), 86–89. Available from: <https://doi.org/10.1038/nature13977>
- Crosetto, P. & Filippin, A. (2013) The ‘bomb’ risk elicitation task. *Journal of Risk and Uncertainty*, 47(1), 31–65. Available from: <https://doi.org/10.1007/s11166-013-9170-z>
- Dai, Z., Galeotti, F. & Villeval, M.C. (2018) Cheating in the lab predicts fraud in the field: an experiment in public transportation. *Management Science*, 64(3), 1081–1100. Available from: <https://doi.org/10.1287/mnsc.2016.2616>
- Dufwenberg, M. & Dufwenberg, M.A. (2018) Lies in disguise—a theoretical analysis of cheating. *Journal of Economic Theory*, 175, 248–264. Available from: <https://doi.org/10.1016/j.jet.2018.01.013>
- Effron, D.A. & Conway, P. (2015) When virtue leads to villainy: advances in research on moral self-licensing. *Current Opinion in Psychology*, 6, 32–35. Available from: <https://doi.org/10.1016/j.copsyc.2015.03.017>
- Engelmann, J.B. & Fehr, E. (2016) The slippery slope of dishonesty. *Nature Neuroscience*, 19(12), 1543–1544. Available from: <https://doi.org/10.1038/nn.4441>
- Ermisch, J. & Gambetta, D. (2006) People’s trust: the design of a survey-based experiment. resreport. IZA Discussion Paper No. 2216.

- Ermisch, J., Gambetta, D., Laurie, H., Siedler, T. & Noah Uhrig, S. (2009) Measuring people's trust. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4), 749–769. Available from: <https://doi.org/10.1111/j.1467-985x.2009.00591.x>
- Falk, A. & Kosfeld, M. (2006) The hidden costs of control. *The American Economic Review*, 96(5), 1611–1630. Available from: <https://doi.org/10.1257/aer.96.5.1611>
- Fischbacher, U. (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178. Available from: <https://doi.org/10.1007/s10683-006-9159-4>
- Fischbacher, U. & Föllmi-Heusi, F. (2013) Lies in disguise - an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547. Available from: <https://doi.org/10.1111/jeea.12014>
- Garbarino, E., Slonim, R. & Villeval, M.C. (2018) A method to estimate mean lying rates and their full distribution. *Journal of the Economic Science Association*, 4(2), 136–150. Available from: <https://doi.org/10.1007/s40881-018-0055-4>
- Garrett, N., Lazzaro, S.C., Ariely, D. & Sharot, T. (2016) The brain adapts to dishonesty. *Nature Neuroscience*, 19(12), 1727–1732. Available from: <https://doi.org/10.1038/nn.4426>
- Geraldes, D., Heinicke, F. & Rosenkranz, S. (2019) Lying on two dimensions and moral spillovers. MPRA Paper No. 96640.
- Gino, F. & Ariely, D. (2012) The dark side of creativity: original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, 102(3), 445–459. Available from: <https://doi.org/10.1037/a0026406>
- Gioia, F. (2016) Peer effects on risk behaviour: the importance of group identity. *Experimental Economics*, 20, 1–30. Available from: <https://doi.org/10.1007/s10683-016-9478-z>
- Gneezy, U., Kajackaite, A. & Sobel, J. (2018) Lying aversion and the size of the lie. *The American Economic Review*, 108(2), 419–453. Available from: <https://doi.org/10.1257/aer.20161553>
- Greiner, B. (2015) Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114–125. Available from: <https://doi.org/10.1007/s40881-015-0004-4>
- Hofmann, W., Wisneski, D.C., Brandt, M.J. & Skitka, L.J. (2014) Morality in everyday life. *Science*, 345(6202), 1340–1343. Available from: <https://doi.org/10.1126/science.1251560>
- Ichino, A. & Muehlheusser, G. (2008) How often should you open the door? Optimal monitoring to screen heterogeneous agents. *Journal of Economic Behavior & Organization*, 67(3–4), 820–831. Available from: <https://doi.org/10.1016/j.jebo.2008.02.004>
- Khalmetski, K. & Sliwka, D. (2019) Disguising lies — image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics*, 11(4), 79–110. Available from: <https://doi.org/10.1257/mic.20170193>
- 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, 102270. Available from: <https://doi.org/10.1016/j.joep.2020.102270>
- Lowes, S., Nunn, N., Robinson, J.A. & Weigel, J.L. (2017) The evolution of culture and institutions: evidence from the Kuba Kingdom. *Econometrica*, 85(4), 1065–1091. Available from: <https://doi.org/10.3982/ecta14139>
- Mann, H., Garcia-Rada, X., Hornuf, L., Tafurt, J. & Ariely, D. (2016) Cut from the same cloth: similarly dishonest individuals across countries. *Journal of Cross-Cultural Psychology*, 47(6), 858–874. Available from: <https://doi.org/10.1177/0022022116648211>
- Mol, J.M., van der Heijden, E.C. & Potters, J.J. (2020) (Not) alone in the world: cheating in the presence of a virtual observer. *Experimental Economics*, 23(4), 961–978. Available from: <https://doi.org/10.1007/s10683-020-09644-0>
- Monin, B. & Miller, D.T. (2001) Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43. Available from: <https://doi.org/10.1037/0022-3514.81.1.33>
- Pittarello, A., Leib, M., Gordon-Hecker, T. & Shalvi, S. (2015) Justifications shape ethical blind spots. *Psychological Science*, 26(6), 794–804. Available from: <https://doi.org/10.1177/0956797615571018>
- Ploner, M. & Regner, T. (2013) Self-image and moral balancing: an experimental analysis. *Journal of Economic Behavior & Organization*, 93, 374–383. Available from: <https://doi.org/10.1016/j.jebo.2013.03.030>
- de Quidt, J., Haushofer, J. & Roth, C. (2018) Measuring and bounding experimenter demand. *The American Economic Review*, 108(11), 3266–3302. Available from: <https://doi.org/10.1257/aer.20171330>
- Sappington, D.E. (1991) Incentives in principal-agent relationships. *The Journal of Economic Perspectives*, 5(2), 45–66. Available from: <https://doi.org/10.1257/jep.5.2.45>
- Shalvi, S. (2012) Dishonestly increasing the likelihood of winning. *Judgment and Decision Making*, 7(3), 292–303. Available from: <https://doi.org/10.1017/s1930297500002266>
- Shalvi, S., Dana, J., Handgraaf, M.J. & De Dreu, C.K. (2011) Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181–190. Available from: <https://doi.org/10.1016/j.obhdp.2011.02.001>
- Suri, S., Goldstein, D.G. & Mason, W.A. (2011) Honesty in an online labor market. *Human Computation*, 11, 61–66.
- Walker, B. (2000) Monitoring and motivation in principal-agent relationships: some issues in the case of local authority services. *Scottish Journal of Political Economy*, 47(5), 525–549. Available from: <https://doi.org/10.1111/1467-9485.00177>
- Whynes, D.K. (1993) Can performance monitoring solve the public services' principal-agent problem? *Scottish Journal of Political Economy*, 40(4), 434–446. Available from: <https://doi.org/10.1111/j.1467-9485.1993.tb00665.x>
- Zizzo, D.J. (2010) Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75–98. Available from: <https://doi.org/10.1007/s10683-009-9230-z>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Casal, S. & Filippin, A. (2023) The effect of observing multiple private information outcomes on the inclination to cheat. *Economic Inquiry*, 1–20. Available from: <https://doi.org/10.1111/ecin.13197>

APPENDIX A

INSTRUCTIONS (TRANSLATION FROM ITALIAN)

Welcome! Thank you for participating in this experiment.

The instructions are displayed on your screen and they will also be read aloud. Please, follow carefully the instructions since your final earning will depend on the decisions taken during the experiment. Thus, it is in your personal interest to have a thorough comprehension of the experiment.

If you have any question at any time please raise your hand and wait for an experimenter to come to your desk and answer it privately. During the whole experiment it is prohibited to talk with other participants.

The experiment starts with a short questionnaire and will then continue with three Phases. We start reading the instructions of Phase 1. You will then receive the instructions for Phase 2 and 3 in due course. In each Phase, you will earn ECU. The earnings of each Phase are independent and you will be paid the sum of your earnings in every Phase. The total amount in ECU will be converted using the exchange rate

$$20 \text{ ECU} = 1 \text{ EURO}$$

and paid privately in cash at the end of the experiment. For your participation you will also receive an additional amount of 2.5 euro.

All the decisions made during the experiment are anonymous. At the end of the experiment we will call you individually using the number randomly drawn when entering the lab and we will give you the total amount earned in cash.

PHASE 1

A grid with 100 boxes will appear on your computer.

Your task is to choose how many boxes to collect. So, you will be asked to choose a number between 1 and 100. Boxes will be collected in numerical order, starting from the box in the top left corner of the grid.

Every box collected is worth 1 ECU. However, this earning is only potential since one of this boxes hides a time bomb. You do not know where this bomb lies. You only know that the bomb can be in any box with equal probability. After choices have been made, the computer will randomly determine which box contains the bomb. This random draw is made at the individual level, thus the box containing the bomb can differ for every participant.

If the bomb is located in a box that you did not collect—that is, the number of boxes you choose is smaller than the number of the box containing the bomb—you will earn 1 ECU for each box collected.

In contrast, if you happen to collect the box where the bomb is located—that is, if the number of boxes you choose is greater than or equal to the number of the box containing the bomb—the bomb by exploding will destroy the earnings: thus, you will earn zero in Phase 1.

PHASE 2

In this phase two participants interact: *Player A* and *Player B*. The figure on your screen (see Figure 2) summarizes all possible outcomes of this game.

At the beginning, *Player A* is endowed with 10 ECU and must decide among two alternatives, called OPTION 1 and OPTION 2.

OPTION 1

Player A keeps the 10 ECU and nothing is passed to *Player B*. The game ends without *Player B* taking any decision and earnings for Phase 2 will be 10 ECU for *Player A* and 0 ECU for *Player B*, respectively.

OPTION 2

Player A passes 10 ECU to *Player B*. The 10 ECU will be multiplied by a factor of 4, thus *Player B* will receive 40 ECU.

In this case *Player B* must decide among **OPTION 3** and **OPTION 4**.

OPTION 3

Player B keeps the 40 ECU and nothing is passed back to *Player A*. Earnings for Phase 2 will be 0 ECU for *Player A* and 40 ECU for *Player B*, respectively.

OPTION 4

Player B keeps the 18 ECU and passes back 22 ECU to *Player A*. Earnings for Phase 2 will be 22 ECU for *Player A* and 18 ECU for *Player B*, respectively.

After the choices have been made, the computer will randomly form groups of two players, randomly assigning a role (*Player A* and *Player B*) to each player. For this reason, you will be asked to make a decision both as *Player A* and as *Player B*. Once pairs have been formed, earnings will be automatically computed according to the decision made by each player in the assigned role.


























PHASE 3


In this phase you are required to recognize the value and the issuing country of several euro coins. As shown in the example of Figure A1 on the left part of your screen, a table with several coins will appear. The corresponding value and the issuing country is reported below each coin. On the right part of your screen you will see one coin randomly selected from the table.

Your task is to identify both the value and the issuing country of the selected coin. You have to select the value and the country from the corresponding list and then confirm your choice. You will then be notified about the correctness of your response. Your task will then continue with the identification of another coin randomly chosen from a different table.

Secondi rimanenti per questa fase: 0:04

Seleziona il paese d'origine e l'importo della moneta in Euro visualizzata

 0.5 Slovenia	 0.1 Città del Vaticano	 0.5 Lituania	 0.1 Germania	 0.2 Spagna
 0.1 San Marino	 0.1 Slovacchia	 0.5 Monaco	 0.2 Città del Vaticano	 0.5 Grecia
 0.1 Irlanda	 0.5 Andorra	 0.5 Finlandia	 0.2 Portogallo	 0.5 Città del Vaticano
 0.1 Francia	 0.5 Cipro	 0.1 Austria	 0.2 Lettonia	 0.2 Belgio
 0.2 Italia	 0.1 Lussemburgo	 0.2 Olanda	 0.2 Estonia	 0.5 Malta



PROSEGUI

Paese	#	Importo	#
Andorra	<input type="radio"/>	0.01€	<input type="radio"/>
Austria	<input type="radio"/>	0.02€	<input type="radio"/>
Belgio	<input type="radio"/>	0.05€	<input type="radio"/>
Cipro	<input type="radio"/>	0.10€	<input type="radio"/>
Estonia	<input type="radio"/>	0.20€	<input type="radio"/>
Finlandia	<input type="radio"/>	0.50€	<input type="radio"/>
Francia	<input type="radio"/>	1€	<input type="radio"/>
Germania	<input type="radio"/>	2€	<input type="radio"/>
Grecia	<input type="radio"/>		
Irlanda	<input type="radio"/>		
Italia	<input type="radio"/>		
Lettonia	<input type="radio"/>		
Lituania	<input type="radio"/>		
Lussemburgo	<input type="radio"/>		

FIGURE A1 Screenshot of the Coin Task.

You have 20 min to identify as many coins as possible, with a maximum of 20. There are 60 tables available, and therefore you can successfully complete the task even if you make some mistakes. At the end of each round you will also be informed about your total score up to that point. Before starting with the real task, you have the opportunity to practice for 60 s. The coins identified during this trial period do not count for your score.

Your earnings in this Phase of the experiment are composed by two parts:

1. A deterministic component linked to your score: you will receive 2 ECU for each coin correctly identified.
2. A variable component unrelated to the number of coins correctly identified.

1-Uniform: This part of your compensation is worth between 10 and 60 ECU, and will be privately determined randomly drawing one number. The number represents a value in ECU (e.g., the number 10 is worth 10 ECU, the number 11 is worth 11 ECU, and so on).

[Wave 1]: On your desk, there is bag containing 51 numbered tokens, from 10 to 60. Draw only one number from this bag and report it on your computer.

[Wave 2]: To generate the number we will use one of the random number generators available on the internet, specifically *calculator.net*. Note that the generator is not operated by us. To ensure anonymous access, we ask you to view the result using your smartphone or a device other than the PC you are currently using for the experiment. *calculator.net* generates the random number as if it was drawn from an urn containing 51 different numbers, specifically all the integers from 10 to 60. The next screen will provide the necessary information to access *calculator.net*.^a After drawing the number, please report it on your computer.

^aThe instructions continued by showing a QR code that directed the participants to <https://www.calculator.net/random-number-generator.html> with specific instructions, including illustrative screen-shots, on how to set up the random number generator correctly.

10-Uniform: This part of your compensation is worth between 10 and 60 ECU, and will be privately determined randomly drawing for 10 times a number between 1 and 6. Each number represents a value in ECU (for instance, number 1 means 1 ECU, number 2 means 2 ECU and so on). The sum of these numbers, which must be by construction between 10 and 60, represents the variable part of your earnings in this phase. If helpful, feel free to use a sheet of paper or your smartphone to write down the numbers drawn and/or to compute the sum.

[Wave 1]: On your desk, there is bag containing 6 numbered tokens, from 1 to 6. Draw one number from this bag for 10 times and report the sum of the 10 draws on your computer. ATTENTION: each draw must be done with all the 6 numbers in the bag. Hence, remember to put the number back into the bag before proceeding with the subsequent draw. Repeat the procedure for exactly 10 times.

[Wave 2]: To generate the number we will use one of the random number generators available on the internet, specifically *calculator.net*. Note that the generator is not operated by us. To ensure anonymous access, we ask you to view the result using your smartphone or a device other than the PC you are currently using for the experiment. *calculator.net* generates the random numbers by rolling a standard 6-sided die, resulting in numbers from 1 to 6. You have to repeat the process until you have the results of 10 die rolls. The next screen will provide the necessary information to access *calculator.net*.^a After drawing the numbers, please report the sum on your computer.

^aThe instructions continued by showing a QR code that directed the participants to <https://www.calculator.net/dice-roller.html> with specific instructions, including illustrative screen-shots, on how to set up the random number generator correctly.

1-Normal: This part of your compensation is worth between 10 and 60 ECU, and will be privately determined randomly drawing one number. The number represents a value in ECU (e.g., the number 10 is worth 10 ECU, the number 11 is worth 11 ECU, and so on).

To generate the number we will use one of the random number generators available on the internet, specifically *pinetools.com*. Note that the generator is not operated by us. To ensure anonymous access, we ask you to view the result using your smartphone or a device other than the PC you are currently using for the experiment. *pinetools.com* generates the random number using a procedure equivalent to rolling 10 times a standard 6-sided die, each resulting in numbers from 1 to 6, and taking the sum. The random generator will show you only the sum of the outcomes obtained. You have to report this number on your computer. The reported number, which must be by construction between 10 and 60, represents the variable part of your earnings in this phase. The next screen will provide the necessary information to access *pinetools.com*.^a

^aThe instructions continued by showing a QR code that directed the participants to <https://pinetools.com/gaussian-random-number-generator> with specific instructions, including illustrative screen-shots, on how to set up the random number generator correctly.

You will start with the identification of the coins, then you will proceed with the determination of the variable part of your earnings.

APPENDIX B

Knowledge of joint probabilities

In a pre-experimental questionnaire we ask participants what is the probability to obtain a sum equal to 2 and to 7 rolling two dice.

Both cases contain the following alternatives:

- a) $1/6$
- b) $1/12$
- c) $1/36$
- d) *I don't know*

When asking about the probability of obtaining a sum equal to 7 the following alternatives are added:

- e) *I don't know exactly, but smaller than obtaining a 2*
- f) *I don't know exactly, but the same as obtaining a 2*
- g) *I don't know exactly, but larger than obtaining a 2*

We define a participant as being unfamiliar with joint probabilities if the answer to the second question is different from a) or g).

Regardless of the correctness of their answers, participants receive the solution of the two questions and a brief explanation of why it is more likely to obtain 7 than 2.