

Caliari, Daniele; Soraperra, Ivan

Working Paper

Planning to cheat: Temptation and self-control

WZB Discussion Paper, No. SP II 2023-205

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Caliari, Daniele; Soraperra, Ivan (2023) : Planning to cheat: Temptation and self-control, WZB Discussion Paper, No. SP II 2023-205, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:
<https://hdl.handle.net/10419/277737>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WZB

Wissenschaftszentrum Berlin
für Sozialforschung



Daniele Caliari
Ivan Soraperra

Planning to cheat. Temptation and self-control

Discussion Paper

SP II 2023–205

September 2023

WZB Berlin Social Science Center

Research Area

Markets and Choice

Research Unit

Market Behavior

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Copyright remains with the author(s).

Discussion papers of the WZB serve to disseminate the research results of work in progress prior to publication to encourage the exchange of ideas and academic debate. Inclusion of a paper in the discussion paper series does not constitute publication and should not limit publication in any other venue. The discussion papers published by the WZB represent the views of the respective author(s) and not of the institute as a whole.

Daniele Caliari, Ivan Soraperra

Planning to cheat. Temptation and self-control

Affiliation of the authors:

Daniele Caliari

WZB Berlin Social Science Center

Ivan Soraperra

Max Planck Institute for Human Development

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

Planning to cheat. Temptation and self-control

by Daniele Caliari and Ivan Soraperra ^{*}

Are opportunities making thieves? Accumulated experimental evidence shows that, when people have the opportunity to cheat, often they take it. Most of the literature on cheating opportunities forces people into a tempting situation where they face a trade-off between money and morality. In our paper, we ask whether people are sophisticated in their cheating behavior and whether they search for or avoid these trade-offs. Overall, participants in the experiment exhibit very little temptation, i.e., virtually no one is willing to pay a cost to avoid the possibility of misreporting in a coin-flip-like task, and they are able to consistently stick to their plan. Participants with a strict preference for the tempting situation, i.e., who are planning to cheat, show a winning rate of about 95% and those that are indifferent between having and not having the opportunity, i.e., who are planning to be honest, show a winning rate that is close to 50%.

Keywords: temptation and self-control, cheating, unethical behavior, lab experiment

JEL classification: B41, C91, C93

^{*}E-mail: daniele.caliari@wzb.eu, soraperra@mpib-berlin.mpg.de.

1 Introduction

[Francis Bacon] "Opportunity makes a thief."

In the opening quote, Francis Bacon suggests that people can be cheaters, but they may be occasional ones. They seize cheating opportunities because resisting the temptation to gain undeserved profits is a challenging task, and occasionally, they succumb to it. However, they may not actively plan to pursue such opportunities. In this paper, we explore whether this view holds true.

The existing body of evidence on lying and cheating offers limited insights into this phenomenon. Existing studies primarily involve placing individuals in tempting situations to observe their responses, measuring the extent to which they cheat or lie when facing the opportunity to cheat (see, e.g., [Fischbacher & Föllmi-Heusi \(2013\)](#), [Utikal & Fischbacher \(2013\)](#), [Gneezy et al. \(2018\)](#), [Kajackaite & Gneezy \(2017\)](#) for some notable examples, and [Abeler et al. \(2019\)](#) and [Gerlach et al. \(2019\)](#) for meta-analytical reviews of the literature). An important yet often overlooked factor is whether people intentionally plan to cheat or only do so when presented with an opportunity.¹ While some individuals may give in to temptation when presented with the chance, they may actively avoid situations where they could be tempted. In this paper, we study whether people plan to cheat, and if they do so, whether they keep up with their plan or not. By considering the role of planning in cheating behavior, we can gain a more nuanced understanding of why and how individuals cheat. This, in turn, can inform the development of effective interventions and policies to prevent cheating in various settings.

To illustrate our point, consider the following example. A shopper at the exit of the supermarket has the choice to either go to an unsupervised self-checkout or to a staffed counter. While the second option offers no opportunity to cheat, the first gives the shopper the possibility to cheat avoiding scanning some items. The choice of the counter can be difficult for the shopper, as it entails a trade-off between personal gain and his/her own moral standards. The shopper must consider not only (i) the potential economic benefits but also (ii) the guilt or loss of self-image associated with not scanning certain items and (iii) the possible cost of resisting the temptation to cheat when using the self-checkout.

Assume at first that temptation does not play a role. A honest person, i.e., a decision-

¹Notable exception is a recent paper by [Konrad et al. \(2021\)](#). See Section 5.5 for a discussion of the differences between ours and their approach.

maker who values their moral standards highly enough, would choose the closest of the two checkouts, regardless of whether this option offers cheating opportunities, and would scan all the items. Conversely, a cheater, i.e., a decision maker who does not experience significant guilt from not scanning items, would always choose the self-checkout, even if it is far away, and would exploit the opportunity to cheat. Here, the cost of walking to the counter determines how cheaters and honest shoppers select themselves: when the cost of reaching the staffed counter is lower than the cost of reaching the self-checkout, honest shoppers and cheaters would go to different counters; when the cost of reaching the staffed counter is higher than the cost of reaching the self-checkout, instead, everyone goes to the self-checkout. As a result, without temptation no-one would choose the staffed counter when this is more costly to reach.

When temptation plays a role, instead, shoppers may experience a cost when facing a tempting situation and they may proactively exert self-control by avoiding the unsupervised checkout. So, honest shoppers who experience significant temptation may be willing to pay a cost and walk to the staffed counter, even when the self-checkout is easier to reach. In this case, some shoppers who might give-in to temptation may avoid the self-checkout altogether to commit to not cheating. Therefore, if temptation plays a role, the selection process is different. We may observe a significant number of shoppers going to the staffed counter even when the cost of reaching it is higher than the cost of reaching the self-checkout. As a result, by manipulating the cost of reaching the different counters, one can infer whether people plan to cheat, whether they stick to the plan after the choice, and whether cheating is indeed a temptation that people prefer to avoid.

In this paper, we experimentally and theoretically study the stylized situation described in the shopper example. To do so, we propose a novel approach that models cheating opportunities as a tempting good, as they have the two characterizing properties of these goods (Toussaert, 2018): (i) they are appealing; in our case in monetary (or probabilistic) terms; (ii) their “consumption” is perceived as bad. Building on the theoretical framework of Gul & Pesendorfer (2001) (hereafter GS model), we assume that individuals possess two utility functions representing temptation and commitment, respectively. The commitment utility represents the long-term utility of the agent making a decision in a cool state without temptation. On the other hand, the temptation utility reflects the cost of resisting temptation when making a decision in a hot state, where the individual is directly

confronted with tempting options. Within this framework, agents want to maximize their long-term utility, which is the utility in the cold state, and therefore may want to limit their set of available options if they present a temptation cost.

Our experimental design is a direct analog of the shopper example. Participants were offered a choice between two lotteries, a **Public** and a **Private** lottery, with both lotteries offering the possibility of winning a prize of 15 euros. In addition to choosing a lottery, participants had to personally fill the bags with white and black balls used to play the lotteries. Similar to the example, in the **Public** case, bag-filling was supervised, so honesty was the only option. In contrast, in the **Private** case, bag-filling was unsupervised, leaving the participant free to fill the bag honestly or dishonestly. In the experiment, we manipulate the relative cost of choosing one lottery over the other by changing the probability of winning the prize in the **Public** lottery. While in the **Private** lottery the nominal probability of winning is 50%, in the **Public** it ranges from 100% to 0% making it more or less expensive to choose the latter over the former.

Our results provide evidence of low levels of temptation on two grounds: (i) very few participants choose the **Public** lottery when it yields a lower return than the **Private** lottery; (ii) the winning share of those choosing the **Public** lottery when it yields the same return as the **Private** lottery and are randomly assigned to play the **Private** lottery is not significantly different from 50%, implying that even in this group virtually no one succumbed to temptation. Moreover, results show that, when facing the tempting situation, participants choosing the **Private** lottery when it yields a strictly lower return cheat significantly more than those choosing the **Public** lottery. This result is remarkably evident, as the winning share in the **Private** lottery for the former is 95% while for the latter is close to 50%. Taken together, our results reject Bacon’s conjecture that “opportunity makes a thief” in favour of the idea that people plan to cheat and consistently execute their plans. Finally, we manipulate both the cheating and temptation cost with a treatment adding a negative externality of 2 euros missed donation to a given charity to the cost of cheating. We find no significant effect of the treatment on the lottery choice points and on the winning shares in the **Private** lottery.

The rest of the paper is organized as follows: in Section 2, we introduce the decision problem faced by our decision makers and discuss the application of the GS model to it; Section 3 explains the design of the experiment, gives the procedural details, and outlines

our hypotheses; Section 4 presents the results; and Sections 5 and 6 discuss the related literature, the implications and limitations, and conclude.

2 Decision problem and model

The problem faced by our decision maker, which is described in Figure 1, has two stages. In the **MPL** (Multiple Price List) stage, the decision maker has to choose between a **Public** lottery and a **Private** lottery. If she chooses the **Private** lottery, she moves to the **LG** (Lottery Game) stage where she has the choice of either being honest or cheating. In the former case, she wins 15 euros with a probability of 0.5, while in the latter case, she wins 15 euros for sure, but incurs a cost for having cheated. If she chooses the **Public** lottery, she wins 15 euros with a probability of p .

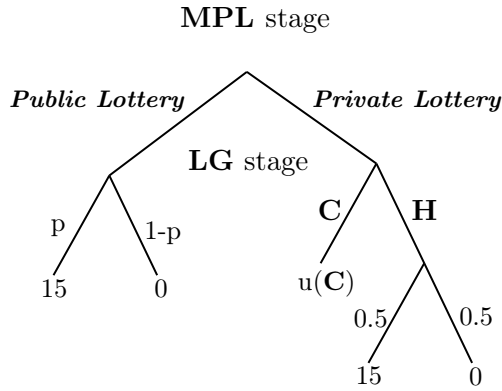


Figure 1: Extensive form representation of the decision problem

The decision maker has, therefore, three options: (i) she can choose the **Private** lottery and cheat (**C**), (ii) she can choose the **Private** lottery and behave honestly (\mathbf{H}_{pr}), (iii) she can behave honestly by choosing the **Public** lottery (\mathbf{H}_{pu}). Following the two-stage formal setup of the GS model, at the **MPL** stage the decision maker has to make a choice between two sets of options: the **Private** lottery, which is represented by the set $\{\mathbf{H}_{pr}, \mathbf{C}\}$, and the **Public** lottery, which is represented by the singleton $\{\mathbf{H}_{pu}\}$. At the **LG** stage, the decision maker has to choose one of the options from the chosen set, which translates to choosing between \mathbf{H}_{pr} and **C** in the former case, and not having a choice in the latter case.

Following GS, we assume that the decision maker's behaviour is governed by two utility functions: the commitment utility u and the temptation utility v . Furthermore, the choice between $\{\mathbf{H}_{pu}\}$ and $\{\mathbf{H}_{pr}, \mathbf{C}\}$ in the **MPL** stage is made in the cold state and the choice

between \mathbf{H}_{pr} and \mathbf{C} in the \mathbf{LG} stage is made in the hot state. Cold and hot states are separated by the role of the temptation utility. On the one hand, in the cold state, the decision maker considers temptation to be costly and so she may try to avoid the most tempting alternative. On the other hand, in the hot state, she is tempted and, therefore, temptation increases the perceived utility of the tempting option. The model operationalizes this intuition by assuming that, during the second stage (hot), the decision maker maximizes $u(x)+v(x)$ when selecting from the set. Conversely, during the first stage (cold), the decision maker chooses anticipating temptation and computes the utility $U(A)$ of each set using equation 1, where the cost of temptation is represented by $\max_{y \in A} v(y)$.

$$U(A) = \max_{x \in A} (u(x) + v(x)) - \max_{y \in A} v(y) \quad (1)$$

In our setup, decision-makers are therefore characterized by a tuple (u, v) as shown in Table 1 below. $u(p)$ is the utility of the **Public** lottery and it depends on the probability of winning $p \in [0, 1]$. Similarly, $u(\mathbf{H}_{pr})$ is the commitment utility of the **Private** lottery when the decision maker behaves honestly, and $u(\mathbf{C})$ is the commitment utility of the decision maker when cheating. Finally, $v(\mathbf{H}_{pr})$ and $v(\mathbf{C})$ are the temptation utilities of behaving honestly and cheating, respectively.²

	u	v
\mathbf{H}_{pu}	$u(p)$	0
\mathbf{H}_{pr}	$u(\mathbf{H}_{pr})$	$v(\mathbf{H}_{pr})$
\mathbf{C}	$u(\mathbf{C})$	$v(\mathbf{C})$

Table 1: Values of the two utility functions u and v for the options in the decision problem

Therefore, when choosing whether to cheat or not in the second stage, the decision maker compares the following utilities: $u(\mathbf{H}_{pr}) + v(\mathbf{H}_{pr})$ and $u(\mathbf{C}) + v(\mathbf{C})$. When choosing between the **Public** and **Private** lotteries in the first stage, instead, she compares the following utilities:

$$U(\{\mathbf{H}_{pr}, \mathbf{C}\}) = \max\{u(\mathbf{H}_{pr}) + v(\mathbf{H}_{pr}), u(\mathbf{C}) + v(\mathbf{C})\} - \max\{v(\mathbf{H}_{pr}), v(\mathbf{C})\} \quad (2)$$

²Here we make the simplifying assumption that $v(p) = 0$. Note that this assumption is inconsequential in our setup. This is because $v(p)$ is only used in the evaluation of the singleton $\{\mathbf{H}_{pu}\}$ and, therefore, it cancels out.

$$U(\{\mathbf{H}_{pu}\}) = u(p) \tag{3}$$

In what follows, we make the assumption that, in the cold state, the utility of not cheating in the **Private** lottery is equivalent to the utility of a **Public** lottery with a probability of $p = 0.5$, i.e., $u(\mathbf{H}_{pr}) = u(0.5)$. The rationale behind this assumption is that the decision-maker remains indifferent between having a commitment device that prevents her from cheating by removing temptation and not having the option to cheat at all.

3 Experimental design and hypotheses

This section describes the experimental design used in our study and outlines our hypotheses regarding the impact of temptation and planning to cheat. Firstly, we will describe the main task and outline the experimental procedures. Next, we will formulate our research hypotheses based on the model in Section 2.

3.1 Experimental design

The experiment is divided into three parts: the preparation, the main task, consisting of the **MPL** stage and the **LG** stage, and, finally, the questionnaire.

3.1.1 The main task

In the **MPL** stage of the main task, participants faced a series of 11 binary choices in which they had to choose between the **Private** lottery and a **Public** lottery which offered a probability p to win the prize that was decreasing in each of the 11 binary choices (see Table 2). As explained, in the **Public** lottery, participants knew that the bag used to play the lottery was prepared under the experimenter’s supervision. In contrast, in the **Private** lottery, participants knew that they would be preparing the bags themselves, without the possibility of supervision by the experimenter. The binary choices in the **MPL** stage were presented in the same order reported in Table 2, but each choice was presented on a *separate screen*. After the participant made her choices in each of the screens, the computer randomly selected one screen to be played and the participants entered the **LG** stage of the main task, where the lottery chosen by the participant in the randomly selected screen was played out.

In the **LG** stage, each participant entered a separate room where they were provided

Choice nr.	P. win in Public	P. win in Private
#1	100%	50%
#2	80%	50%
#3	60%	50%
#4	55%	50%
#5	50%	50%
#6	45%	50%
#7	40%	50%
#8	30%	50%
#9	20%	50%
#10	10%	50%
#11	0%	50%

Table 2: Multiple price list planning to cheat

with the necessary materials to prepare the bags required for the lottery game.³ If the lottery chosen in the selected screen was the **Public** lottery, both the experimenter and referee would remain in the room with the participant. The participant would then prepare the bag based on the winning probability of the selected lottery. After the participant inserted the correct number of winning (white) and losing (black) balls, the referee drew one ball from the bag. If the ball was white, the participant obtained the prize of 15 euros otherwise they obtained nothing.

If the lottery chosen in the selected choice was the **Private**, the experimenter and the referee would leave the participant alone and unobserved in the room. There, on a table, the participant would find two bags and three balls, one losing ball (black) and two winning balls (white). The task was for the participant to put one ball in each of the bags and discard the third ball by throwing it in a container. Once the participant filled the bags, the experimenter and referee were let back into the room, and the referee had to choose one of the two bags. The chosen bag was then opened, and if the ball in the bag was white, the participant won 15 euros. After showing the ball to the experimenter, the referee put the ball back into the bag, and both the referee and experimenter left the

³To ensure a fair and transparent procedure and increase credibility, we followed the methodology proposed by [Baillon et al. \(2022\)](#). At the beginning of the experiment, we randomly selected one participant to act as the referee. The referee’s role was to supervise the random draws in the lottery game and guarantee that the experimenter had no possibility of observing the participant’s decision in the **Private** lottery. The referee was randomly selected at the beginning of each session, prior to reading the instructions. Referees did not make any choice during the experiment and received a fixed payment of 18 euros for their participation.

room. The participant would then insert the two balls in the container with the third ball, ensuring the ball contained in the other bag could not be identified. The **LG** stage was then concluded.

3.1.2 Experimental procedures, demonstration of the task, and the Pledge of Honesty

We conducted the experiment between February and March 2023 at the Technical University of Berlin. The experiment was programmed in zTree (Utikal & Fischbacher, 2013) and participants were recruited via ORSEE (Greiner, 2015). The design and the main hypotheses were pre-registered on as-predicted, n. 121475.⁴ Overall, 196 participants took part in the experiment (14 sessions of 14 participants each). Upon arrival, participants were randomly assigned to a cubicle and the referee was randomly selected.

In the preparation stage before the main task, the experimenter read the instructions aloud and explained how the lotteries in the **LG** stage worked.⁵ Participants were brought to a separate room and the procedure used to fill the bags was demonstrated. This preparation stage was implemented to ensure that the participants had a good understanding of how filling the bags in the two options works and could therefore anticipate temptation. After the demonstration, participants returned to their cubicles and were required to agree to and sign a Pledge of Honesty to proceed with the experiment. The Pledge of Honesty required participants to declare that they would play the **Private** lottery honestly and fill the bag according to the nominal probability of winning.⁶ We made this design choice to increase the moral cost of misreporting and to attenuate the risk that participants would consider cheating opportunities in our experiment as a *rule of the game*. The use of oaths or pledges in experiments has been documented to increase the moral cost of breaking rules and to reduce cheating behavior Jacquemet et al. (2019); Heinicke et al. (2019).

Once participants signed the pledge, they were given time to review the instructions and to answer a series of control questions to ensure they understood the experimental design. Only after correctly answering these questions participants could begin the **MPL** stage of the main task. When all participants completed this stage, they were randomly selected one-by-one to play the **LG** stage in a separate room. After that, they returned to their

⁴The pre-registration can be found here.

⁵Experimental instructions can be found in appendix A.

⁶For the exact wording of the Pledge of Honesty, please refer to the appendix A.

cubicles and completed the final questionnaire, which gathered demographic information such as age, gender, nationality, language, and field of study, as well as assessments of the Big 6 personality traits and of cognitive abilities using standard Raven matrices.

After the questionnaire participants were paid in cash in private and could leave the lab. On average, an experimental session lasted approximately 1 hour and participants earned a payoff of about 18.5 euros each including a show-up fee of 7 euros.

3.1.3 Treatments

We implemented two treatments that differ in the introduction of a negative externality associated with cheating behavior: **no-charity** and **charity** treatment. The **no-charity** treatment follows the above experimental procedure. The **charity** treatment introduces a 2 euros donation to “Save the Children” that is canceled in case of cheating. To implement this, we modify the consequences in the **Public** and **Private** lotteries as follows. On one hand, irrespective of the result of the **Public** lottery the charity receives two euros. On the other hand, if the **Private** lottery is played, the third ball is not discarded, but inserted into a locked box which keys have been given to the referee in the preparation phase. At the end of the experiment, the locked box is opened by the referee and the experimenter. The number of white balls reveals how many subjects have respected the Pledge of Honesty. For each white ball, the charity receives 2 euros. The presence of the charity donation has three main features: (i) it allows us to increase the cost of cheating; (ii) it does not affect the behaviour of the honest participants because they can assure the charity donation behaving truthfully in the **Private** lottery; (iii) it does not affect the observability of cheating behaviour which assures the absence of reputation concerns.

3.2 Hypotheses

We use the GS model to make predictions about the participants’ behavior in our task. The model allows us to make clear-cut predictions with regard to two main variables: (i) the switching point, which is defined by the probability of winning p in the **Public** lottery at which a decision maker starts to choose the **Private** over the **Public** lottery; and (ii) the winning-share in the **Private** lottery conditional on the switching point.

Our predictions rely only on three weak assumptions: (i) $u(0.5) = u(\mathbf{H}_{pr})$ when $p = 0.5$; (ii) monotonicity of $u(p)$ with respect to p ; and (iii) $v(\mathbf{C}) \geq v(\mathbf{H}_{pr})$. The first assumption,

discussed in Section 2, implies that the decision maker is indifferent between having a commitment device preventing her from cheating and not having the option to cheat at all. The second assumption makes the mild requirement that preferences over lotteries respect first-order stochastic dominance.⁷ Importantly, the monotonicity of $u(p)$ implies the uniqueness of a switching point and the direction of the choices should always be from the **Public** to the **Private** lottery. The third assumption implies that the temptation to cheat is higher than the temptation to be honest, which is in line with Bacon’s quote and it is the most natural direction of temptation when applied to cheating behaviour.

Using equations (2) and (3), we can make behavioral predictions based on different values of $u(\mathbf{C})$, $v(\mathbf{C})$, and $v(\mathbf{H}_{pr})$, allowing us to identify four distinct types of participants. Figure 2 summarizes the four types based on the values of u and v , where $u(\mathbf{C})$ is depicted on the y-axis, and the difference between $v(\mathbf{C})$ and $v(\mathbf{H}_{pr})$ is depicted on the x-axis. Type 1 are the “cheaters” because they prefer to cheat even in the cold state when temptation is not present ($u(\mathbf{C}) \geq u(0.5) = u(\mathbf{H}_{pr})$). These participants switch from the **Public** to the **Private** lottery for some $p \geq 0.5$ and, when facing the opportunity to cheat in the hot state, they take it.⁸ Type 2 are the “honest participants that give-in” to temptation. These participants prefer not to cheat in the cold state when temptation is not present ($u(\mathbf{C}) \leq u(0.5) = u(\mathbf{H}_{pr})$), are tempted to cheat ($v(\mathbf{C}) > v(\mathbf{H}_{pr})$), and cannot resist temptation in the hot state ($u(\mathbf{C}) + v(\mathbf{C}) > u(0.5) + v(\mathbf{H}_{pr})$). This type switches to the **Private** lottery when $p < 0.5$ to avoid the tempting situation but when facing temptation the type succumbs and cheats. Type 3 are “honest participants that resist” temptation. These participants prefer not to cheat in the cold state when temptation is not present ($u(\mathbf{C}) < u(0.5) = u(\mathbf{H}_{pr})$), are tempted to cheat ($v(\mathbf{C}) > v(\mathbf{H}_{pr})$), and resist temptation in the hot state ($u(\mathbf{C}) + v(\mathbf{C}) < u(0.5) + v(\mathbf{H}_{pr})$). This type chooses the **Public** lottery when $p < 0.5$ to avoid the temptation cost and, when facing the **Private** lottery, the type behaves honestly. Finally, Type 4 are “honest participants without temptation”. These participants prefer not to cheat in the cold state when temptation is not present ($u(\mathbf{C}) < u(0.5) = u(\mathbf{H}_{pr})$) and do not experience temptation to cheat ($v(\mathbf{C}) = v(\mathbf{H}_{pr})$). This type is indifferent between the **Private** and **Public** lottery when $p = 0.5$ and, since they do not have to avoid temptation, they choose the **Public** lottery if $p > 0.5$ and the **Private**

⁷Note that for our predictions to be valid we don’t require expected utility to hold. Moreover, we don’t make any assumption on the specific shape of cheating costs in the cold state, $u(\mathbf{C})$.

⁸Note that, since both $v()$ and $u()$ rank the alternatives in the same way, for Type 1 temptation does not play a role.

lottery if $p < 0.5$. When facing the latter, they behave honestly.

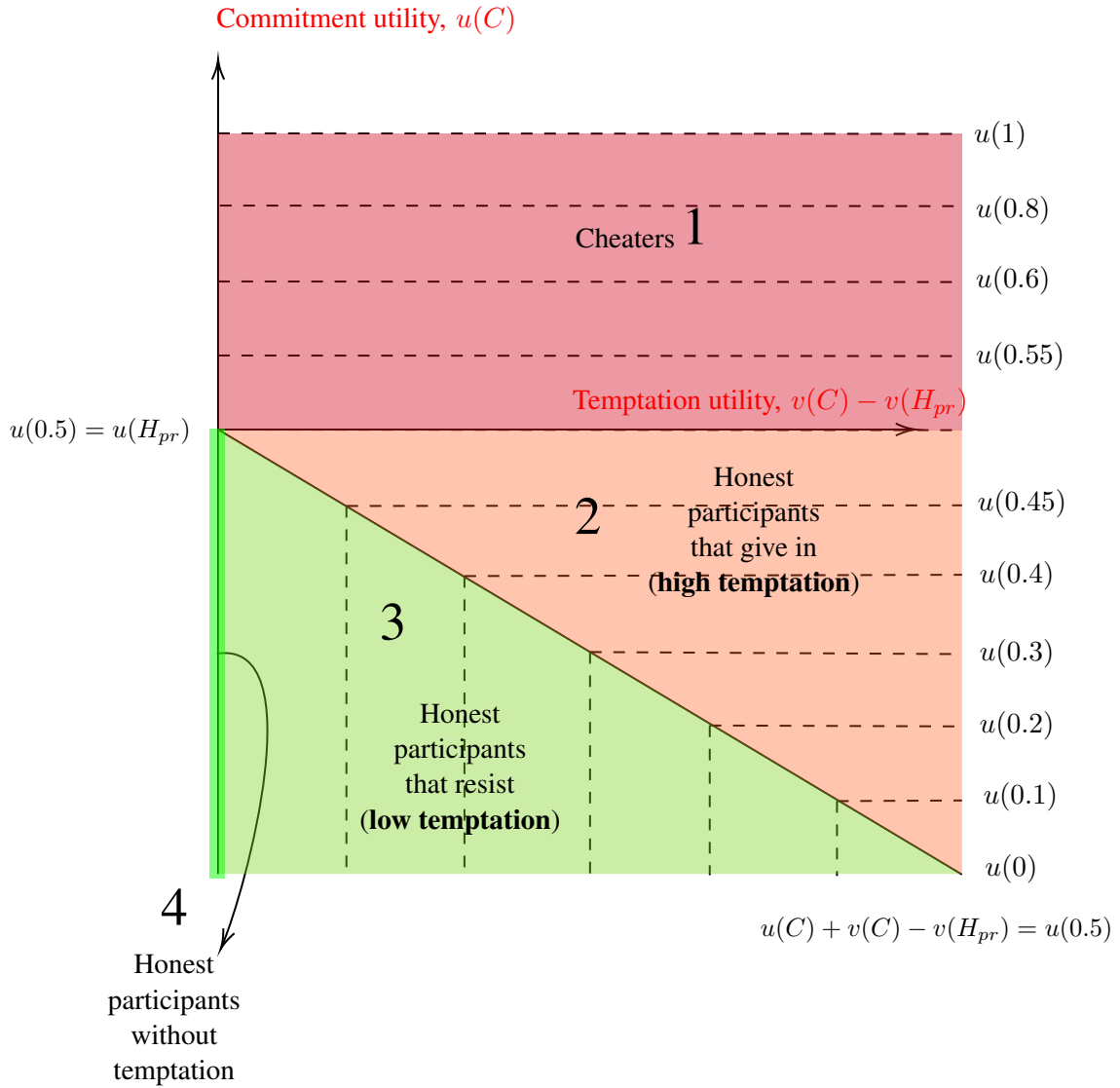


Figure 2: Types when $v(\mathbf{C}) \geq v(\mathbf{H})$

Notes: The dashed lines represent indifference curves for the different levels of p . Whenever a type (u, v) is southeast of a dashed line $u(p)$ she prefers the **Public** lottery characterized by the probability p to the **Private** lottery. The colour of the area represent the behaviour in the **LG** stage with red being cheating and green being honest behaviour.

The types identified in the previous paragraph and represented in Figure 2 lead to the following predictions (independently of the treatments).

Prediction 1 (Temptation) *If temptation is present, a non-negligible fraction of participants switch to the **Private** lottery when $p < 0.45$.*

Prediction 2 (Winning-shares)

1. Conditional on having switched when $p > 0.5$, the winning share in the **Private** lottery is 100%.
2. Conditional on having switched when $p = 0.5$, the winning share in the **Private** lottery is between 50% and 100%.
3. Conditional on having switched when $p < 0.5$ the winning share in the **Private** lottery is between 50% and 100%. Further, in the absence of temptation, the winning share is 50%.

In our **charity** treatment, we modify the commitment and temptation utilities by introducing a negative externality towards the charity. In doing so, we expect the cost of cheating to increase, therefore, reducing $u(\mathbf{C})$. The implications can be seen in Figure 2 as a movement south of the distribution of types. The following predictions arise:

Prediction 3 (Switching points) *On average, participants switch to the **Private** lottery for lower values of p in the **charity** treatment.*

Prediction 3 is driven by the behaviour of Types 1 and 2. For these participants, a movement south in Figure 2 implies, potentially, a lower switching point (horizontal indifference curves). Instead, Types 3 and 4 are unaffected by an increased cost of cheating (vertical indifference curves).

Prediction 4 (Winning-shares) *In the **charity** treatment,*

1. *the winning share in the **Private** lottery for those decision-maker who switch when $p > 0.5$ does not change compared to the **no-charity** treatment;*
2. *the winning share in the **Private** lottery for those decision-maker who switch when $p \leq 0.5$ moves ambiguously compared to the **no-charity** treatment.*

Prediction 4.1 is straightforward, as shown by the color red characterizing Type 1 in Figure 2, the ones switching for $p > 0.5$ still prefer to cheat in the cold state after $u(\mathbf{C})$ decreases. To explain prediction 4.2, notice that since $u(\mathbf{C})$ decreases the number of Type 1 in the region between $u(0.55)$ and $u(0.5)$ may increase/decrease compared to the number of Type 2 implying a higher/lower level of cheating for those who switched when $p = 0.5$. Similarly, a lower cheating cost implies that Type 2 may become Type 3,

therefore reducing the number of participants who succumb to temptation; however, the influx of (tempted) Type 1 becoming Type 2 may increase the number of participants who succumb to temptation.

4 Results

We begin by showing summary statistics in the two treatments: **no-charity** and **charity**. We find that the randomization between treatments has been successful in terms of gender, age, nationality, cognitive abilities, and personality traits, while only field of study seems to significantly differ across treatments (See Table 3).

	Total	charity	no-charity	p-value
Sample size	196	98	98	
Male	64,80%	64,29%	65,31%	0,88
Age	26,65 (7,45)	26,20 (6,92)	27,09 (7,96)	0,41
German (nationality)	70,41%	68,37%	72,45%	0,53
Raven scores	5,76 (1,38)	5,59 (1,30)	5,93 (1,45)	0,089
Field of study				0,034
Economics	28,57%	36,37%	20,41%	
(Other) Social Sciences	19,39%	14,29%	24,49%	
STEM	51,53%	47,96%	55,10%	
Missing	0,51%	1,02%	0,00%	
Personality traits				
Honesty	63,44 (14,52)	61,81 (14,70)	65,08 (14,23)	0,12
Agreeableness	56,11 (14,65)	55,20 (13,54)	57,02 (15,71)	0,39
Extraversion	64,92 (14,66)	64,82 (14,05)	65,03 (15,33)	0,92
Resiliency	55,26 (17,31)	54,95 (15,33)	55,56 (19,17)	0,81
Originality	62,50 (13,14)	62,37 (12,84)	62,63 (13,50)	0,89
Conscientiousness	60,40 (13,63)	61,96 (13,56)	58,83 (13,59)	0,11

Table 3: Descriptive statistics

4.1 The distribution of Types

Figure 3, reports the distributions of switching points in the two treatments. As a first result, we observe that only a small fraction of participants is willing to pay a price in order to avoid the tempting situation. Indeed, only 5% of the participants are switching to the **Private** lottery for $p < 0.45$ — 2 out of 89 in the **no-charity** and 8 out of 94 in the **charity** treatment, respectively —. Contrary to Prediction 1 and Bacon’s opening quote,

this result provides very little evidence in favor of temptation being an important factor for the decision to cheat.

Looking at the effect of externality on the switching point, we find no significant difference between the distributions in the two treatments (two-sample Kolmogorov-Smirnov test, p -value = 0.0626).⁹ Therefore, the externality does not seem to change the cost of cheating in our experiment, providing evidence against Prediction 3.¹⁰ This becomes even clearer when we look at the distributions conditional on switching when $p > 0.45$, which are the ones for which the switching point should change (two-sample Kolmogorov-Smirnov test, p -value = 0.0556). Since some demographics are not perfectly randomized between the two treatments, Table 4 presents regression results testing the effect of introducing the externality on the switching points while controlling for demographics as well as personality traits. Results do not reject the null hypothesis of no effect of the treatment.

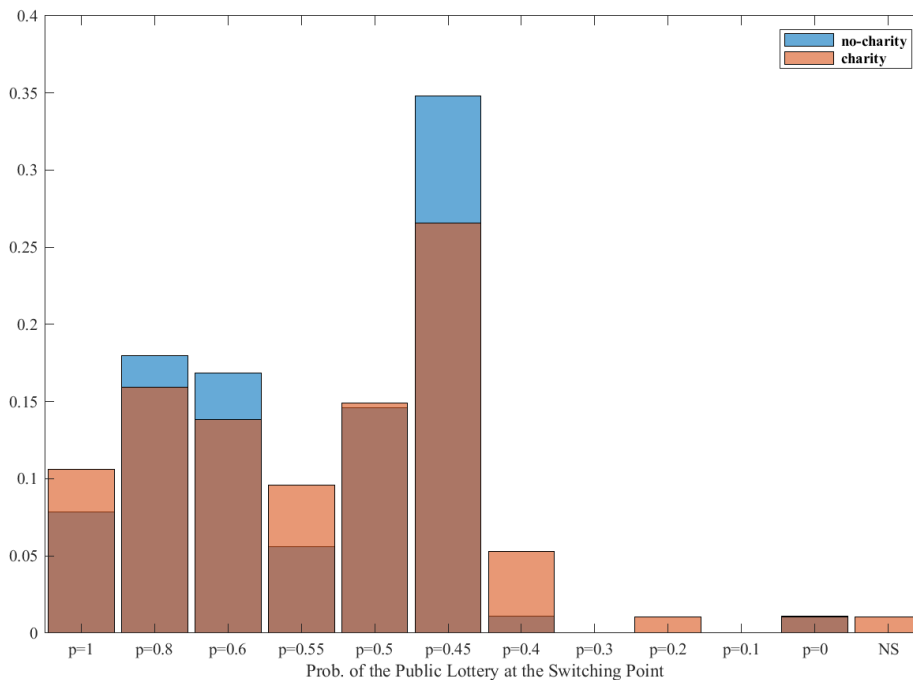


Figure 3: Distribution of switching points by treatment.

⁹The evidence on the effect of externalities on lying is mixed. Gneezy & Kajackaite (2020) find no effect of externalities when stakes are high. Similar results are found in different contexts by Fischbacher & Föllmi-Heusi (2013) and Kajackaite & Gneezy (2017). On the other hand, Sutter (2009) and Erat & Gneezy (2012) find negative effects on lying.

¹⁰As pre-registered, we focus on participants with a unique switching point. Remarkably, only 6.6% of our participants display multiple switching behaviour. We interpret this data as confirmation of the clarity of our instructions.

VARIABLES	(1) Switching Point	(2) Switching Point	(3) Switching Point
Charity	-0.00153 (0.0281)	-0.0130 (0.0303)	-0.0125 (0.0292)
Constant	0.589*** (0.0234)	0.653*** (0.102)	0.785*** (0.144)
Demographic Controls	✖	✓	✓
Personality Traits	✖	✖	✓
Observations	183	183	183
R-squared	0.000	0.034	0.096

Notes: The dependent variable is defined as the probability of winning the **Public** lottery when the **Private** lottery was firstly chosen ("Switching Point"). The variable Charity is our treatment dummy. The demographic controls are gender, field of study, Raven's scores, nationality, and first language (in these cases, we used dummies that take value 1 for German). The personality traits are honesty, agreeableness, extraversion, resiliency, originality, and conscientiousness. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, * <0.1.

Table 4: Treatment effect.

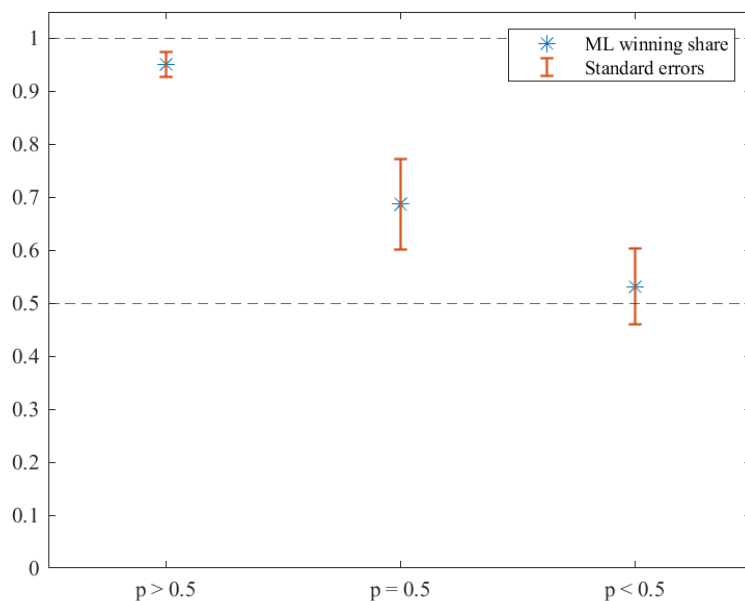
4.2 The winning shares in the Private lottery

Moving to the behavior in the **LG** stage, we look at the share of participants winning the prize in the **Private** lottery conditional on the switching points. Figure 4 shows these shares when collapsing data over treatments. Results closely follows what outlined in Prediction 2: the winning share of the participants switching to **Private** when $p > 0.5$ is 95% (Prediction 2.2); the winning share of the participants switching to **Private** when $p = 0.5$ is 69% (Prediction 2.2); and the winning share of the participants switching to **Private** when $p < 0.5$ is 53% (Prediction 2.3). To have a better idea of the prevalence of cheating, we can estimate the proportions of participants who cheated (p_c) from the winning shares (w) by solving the following equation: $w = p_c + \frac{1}{2}(1 - p_c)$ (Houser et al., 2012). We find that the estimated proportion of cheaters p_c is equal to 90%, 38%, and 6% when $p > 0.5$, $p = 0.5$, and $p < 0.5$, respectively.

Overall, the conditional winning shares demonstrate a remarkable degree of consistency

among our participants. The choices made in the **LG** stage align closely with the predictions derived from their behavior in the **MPL** stage. This finding rejects the idea that participants select the lottery without anticipating temptation or without a clear idea of their future behavior. Instead, it supports the idea that participants plan to either cheat or remain honest, and they are able to adhere to their chosen course of action.

It is worth noting that the winning share among the participants showing temptation, i.e., those who switch for $p < 0.5$, is very close to 50%. This reinforces the notion that temptation does not play a decisive role in guiding the choices of our participants.



Notes: The figure reports the winning shares by switching points (maximum likelihood estimates) as well as the standard errors.

Figure 4: Probability of winning the **Private** lottery.

Looking at the conditional winning shares by treatment (Figure 5). We observe a remarkably consistent behaviour for those participants who switched when $p > 0.5$. A Fisher exact test does not reject the null hypothesis that the winning share is the same in the two treatments (p-value = 0.63), confirming Prediction 4.1. As for the winning shares when $p \leq 0.5$, predictions are ambiguous (Prediction 4.2). Winning shares can either increase or decrease depending on the movement of the different types of participants. Since we found no supporting evidence of a shift in the distribution of the switching points, we expect these shares not to differ. Indeed, we find no difference in the winning share

for participants who switched when $p = 0.5$ (Fisher exact test p-value = 0.30) nor for participants who switched at $p < 0.5$ (Fisher exact test p-value = 1.00).

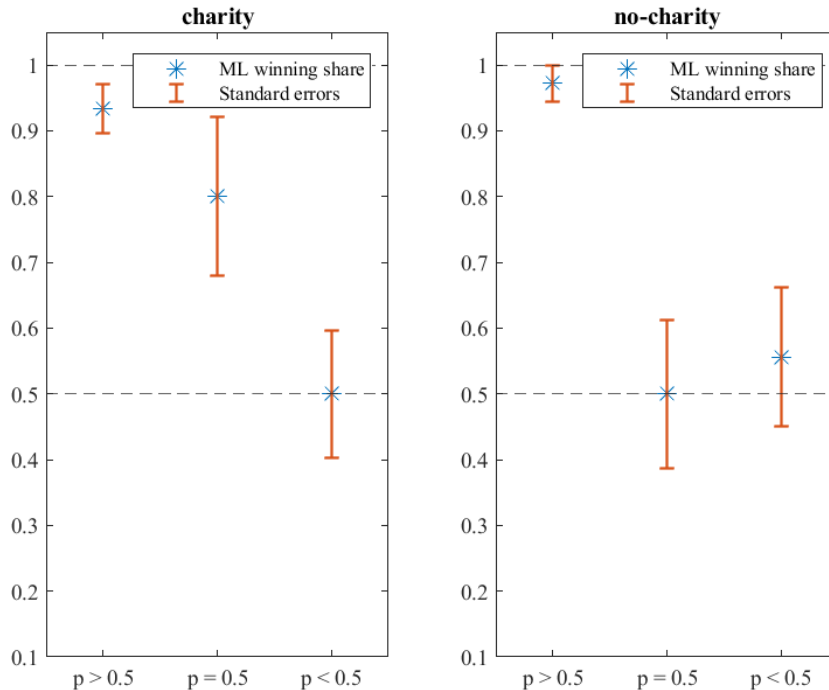


Figure 5: Probability of winning the **Private** lottery by treatment.

4.3 Exploratory analysis: response times

Suggestive evidence of the robustness of our results emerges from analyzing the response times. Examining the time taken by participants to make choices on each screen, we observe that individuals spend significantly more time when they are near their switching point, regardless of their preferences. Figure 6, reports the time spent making a choice as a function of the probability of winning in the **Public** lottery for participants switching above, at, and below $p = 0.5$. Disregarding the time spent on the first screen ($p = 1$), where participants need to understand the interface, we observe that the participants switching above 0.5 take longer to decide when $p = 0.8$, the ones switching at $p = 0.5$ take longer when $p = 0.55$, and those switching below 0.5 take longer when $p = 0.5$ and $p = 0.45$.

These results are in line with the prediction of drift-diffusion and Fechnerian models (see [Ratcliff \(1978\)](#), [Alós-Ferrer et al. \(2021\)](#)) which imply that response times are proportional to utility differences and, therefore, can signal the difficulty of the choice for the decision maker. In our case, response times highlight that people find the choice between **Public**

and **Private** lotteries difficult in a way that is consistent with their type. Moreover, the fact that all participants seem to choose very quickly when $p < 0.45$ suggests once again that temptation plays a minor role in determining the participants' decisions.

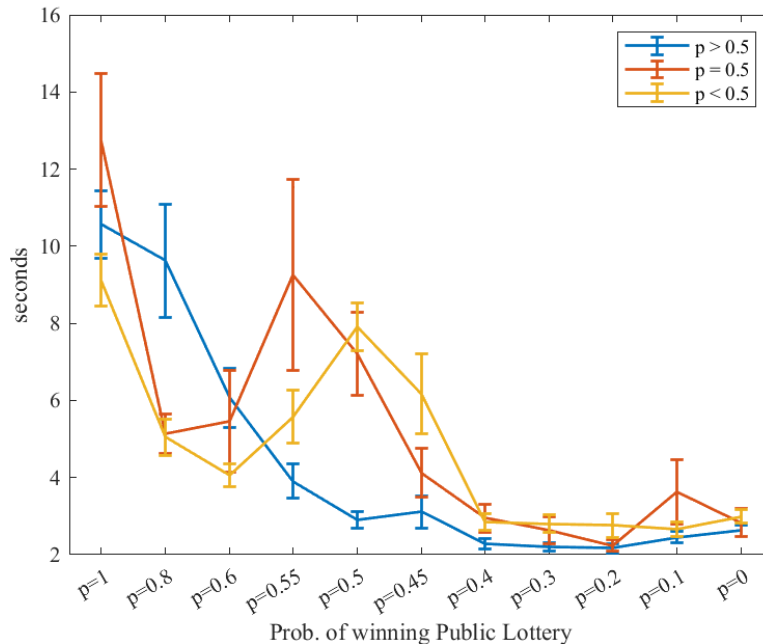


Figure 6: Response Times by type.

The pattern of the response times suggests also that participants think about each problem *separately*, supporting the idea that choices in different screens are made when looking at the options. If the participants had worked out all the choices during the instructions, we should not find significant differences in response times by switching points.

5 Discussion and related Literature

Our discussion is organized as follows. First, we compare our results to the results in the literature studying self-selection into cheating opportunities and evidence for commitment in moral decision-making. Additionally, we extend the discussion to the literature studying the self-selection of workers and managers into organizations and sectors matching their personal moral standards. In doing so, we discuss the issue of the external validity of our stylized setup. Then, we shift the focus to the discussion of the novelty of our experimental design by comparing our contribution to the more standard tasks used to measure cheating in the experimental literature. In doing so, we discuss the merits and the limitations of

our design and potential avenues for further investigation. Finally, we compare our task to previous experiments testing [Gul & Pesendorfer \(2001\)](#)'s model and discuss the interpretation of our results through the lenses of alternative theories of dynamic consistency and preferences for commitment or flexibility.

Sorting into cheating opportunities and demand for moral commitment have received recent attention in the literature ([Shalvi et al., 2011](#); [Konrad et al., 2021](#); [Saccardo & Serra-Garcia, 2023](#)). These papers show that people are willing to pay a price not to face the opportunity to cheat ([Shalvi et al., 2011](#)) and that the people willing to pay a higher price are less likely to take the opportunity to cheat ([Konrad et al., 2021](#)). Our paper contributes to this body of knowledge by unveiling that sorting is not driven by anticipated temptation but is mainly due to perfect planning and heterogeneous cheating costs.¹¹

As for demand for moral commitment, [Saccardo & Serra-Garcia \(2023\)](#) provides evidence of heterogeneous preferences toward moral commitment when individuals are confronted with the trade-off between impartiality and personal gain. Their findings reveal that approximately 40% of the participants seek information to mitigate their self-serving cognitive flexibility, even when such actions come at a cost. Our experiment shows that such a demand for moral commitment does not extend to cheating opportunities. Indeed, we find limited evidence of a willingness to pay for commitment, even when the own profits have external consequences for a charity.

This result relates also to the literature investigating sorting into industries and sectors that are perceived to be more corrupt or immoral ([Banerjee et al., 2015](#); [Hanna & Wang, 2017](#); [Barfort et al., 2019](#)). These studies provide evidence of sorting and of wage premiums associated with positions in morally ambiguous work environments ([Barfort et al., 2019](#); [Schneider et al., 2020](#)), suggesting that people are ready to accept lower wages to work within sectors that exhibit lower levels of corruption or greater moral alignment. One plausible underlying mechanism behind this gap could be the willingness to avoid temptation, with honest individuals ready to pay a premium to avoid the tempting opportunities to cut corners or act immorally. Our results, however, suggest that this mechanism is not a major driving force of the wage premium and points to the potential influence of other factors, such as social pressure, the inability to avoid the immoral act, and reputation — which are

¹¹Note that [Shalvi et al. \(2011\)](#) cannot identify the effect of sorting, since it does not record the choices of those that opt out when facing the cheating opportunity. On the other hand, [Konrad et al. \(2021\)](#) cannot disentangle self-selection due to perfect planning from the effect of temptation, because participants do not have the option to opt-out completely from the tempting situation.

absent in our design. Nevertheless, it is crucial to acknowledge the potential limitations in the external validity of our findings. It is plausible that the avoidance of temptation plays a significant role in driving such wage premiums and that our stylized design might not induce enough temptation and therefore, might not fully capture this aspect. Further research should investigate the roles of temptation avoidance in sorting into moral jobs to clarify this point.

Focusing on the novelty of our experimental design, we discuss the differences with respect to the designs commonly used in literature, highlighting the advantages of our approach. First and foremost, in our design participants commit to either honesty or cheating *before* the realization of the **Private** lottery. This aspect stands in stark contrast to the prevalent experimental designs ([Utikal & Fischbacher \(2013\)](#), [Buccioli & Piovesan \(2011\)](#), [Houser et al. \(2012\)](#), [Gneezy et al. \(2018\)](#), [Konrad et al. \(2021\)](#)), where the decision to engage in dishonest behavior is made only subsequent to the revelation of the actual state of the world. This difference extends to how cheating costs are modeled ([Abeler et al., 2019](#)), rendering our methodology distinct in this regard.

While unconditional commitment may be simply seen as a different way to measure the prevalence of cheating, it offers a practical advantage over the traditional approach. It allows us to estimate the proportion of cheaters based on the distribution of winning shares without the assumption that participants refrain from acting dishonestly against their own self-interest. Indeed, since participants cannot fill the bags with two losing balls, they cannot commit to being dishonest against their self-interest. We believe that this design feature is an appealing feature for future research, especially in scenarios involving trade-offs between personal interests and the welfare of others. This is particularly relevant in situations where the assumption of refraining from lying against one's own self-interest is less likely to be upheld (see [Soraperra et al., 2019](#), for a discussion). Our design choice, however, comes with some logistic costs, as it is not easy to extend to online experiments. The previous advantage, as well as the following ones, must therefore be weighted against the difficulty of implementation.

From the theoretical point of view, the models analysing the traditional design assume that the cost of cheating arises only in the negative scenario of losing.¹² Our setup and modelling choice permits us to sidestep this aspect and, more importantly, does not require

¹²For an in-depth analysis of the distinctions between our design and the more customary designs, see details in Appendix C.

us to make assumptions regarding how decision-makers aggregate utilities across states of the world, be it through expected utility, rank-dependent utility, or other models. These advantages emerge when participants are asked to trade off cheating opportunities for money, a methodology commonly employed when eliciting the Willingness to Pay (WTP) for a **Private** lottery (Shalvi et al., 2011; Konrad et al., 2021). In these cases, accurate measurement of the cost of cheating requires strong assumptions on risk preferences (risk neutrality), on how utility is aggregated across states of the world (expected utility), and on the correlation between risk preferences and cheating costs. All assumptions that are circumvented in our design and modelling choice.¹³

In relation to this last point, our paper contributes to the measurement of cheating costs through the novel channel represented by the opportunity cost of choosing the **Public** lottery as a commitment device. This is a subtle way to measure heterogeneity in cheating costs.¹⁴ As Figure 3 shows, approximately 50% of the participants choose as “cheaters” in the **MPL** stage, but about half of them are ready to give up on at least 20 percentage points of winning chances in order to avoid cheating, i.e., they switch to the **Private** lottery for $p < 0.8$. Overall, considering that a portion of those who make the switch at $p = 0.5$ are also cheaters, we obtain a conservative estimate that 59% of the cheaters in our experiment reveal having a non-negligible cost of cheating. This finding is in line with previous results in the literature showing that a substantial proportion of cheaters have non-zero cheating costs (Gibson et al. (2013), Konrad et al. (2021)). Importantly, Figure 3 shows substantial heterogeneity; a result that differs from Konrad et al. (2021) finding that the majority of the dishonest subjects have a zero cheating cost.

Moving to empirical tests of the model by Gul & Pesendorfer (2001), the closest paper to ours is by Toussaert (2018), which experimentally studies temptation to read a story while performing a tedious task. Contrary to our results, Toussaert (2018) documents that about a quarter to a third of subjects are willing to exert self-control to restrict their choices when this tempting alternative is available. Her paper highlights an important feature, shared by our design, which sets apart the predictions of models of menu-dependent

¹³As an example, Konrad et al. (2021)’s measure of cheating cost (Figure 6 in the paper) assumes risk neutrality, expected utility, and independence of the cheating cost from the probability of winning. Under these assumptions, all participants having a $WTP < 1.5$ and ending up with the bad lottery should cheat. The observed cheating rate of 73% for these participants (Figure 5 in the paper) suggests for misclassification of participants. Indeed, risk-seeking honest participants or honest participants that weight probabilities would be classified as cheaters.

¹⁴We borrow the word “subtle” from Konrad et al. (2021) because it describes very nicely both their as well as our new way of measuring cheating costs.

preferences, such as Gul & Pesendorfer (2001), and the models of dynamically inconsistent preferences (O’Donoghue & Rabin (1999), Frederick et al. (2002)). A unique feature of models of menu-dependent preferences are the “self-control types”, i.e, participants that are ready to restrict their choice to avoid temptation, but are actually able to resist choosing the tempting option when facing it — see Toussaert (2018) for a thorough discussion. Identifying these self-control types requires observing not only participants’ willingness to pay for choice restriction but also their choices when facing the tempting option.¹⁵ Our multiple price list design inherently incorporates this randomization factor, exposing participants to the possibility of encountering the **Private** lottery even if they are willing to pay to avoid temptation. This would have allowed us to identify such types if they had been present. In comparing our approach with Toussaert (2018)’s, it is important to acknowledge an important limitation: our elicitation of menu preferences is forcibly not complete as we do not offer the possibility to commit to cheating (the choice of $\{\mathbf{C}\}$ is not available). Therefore, we cannot identify the willingness to restrict the menu to avoid the temptation to be honest.

To conclude, we interpret our results through the lenses of models capturing preferences for flexibility (Kreps, 1979; Dekel et al., 1998, 2001). In his seminal paper, Kreps (1979) assumes decision makers are uncertain over their own future preferences and that uncertainty is only resolved at a second stage when the state of the world realizes. In this model, decision-makers prefer to keep options open in order to be able to maximize utility in each state of the world. On the one hand, Kreps (1979)’s model allows us to derive predictions on how decision-makers who are uncertain about their type and have a preference for flexibility would behave in our experiment when $p > 0.5$. In this case, the decision-maker has three possible (strict) utility functions ordering the options: $u_1(\mathbf{C}) > u_1(p) > u_1(\mathbf{H}_{pr})$, $u_2(p) > u_2(\mathbf{C}) > u_2(\mathbf{H}_{pr})$, and $u_3(p) > u_3(\mathbf{H}_{pr}) > u_3(\mathbf{C})$. Uncertainty about u_1, u_2, u_3 may lead the decision-maker to choose the **Private** lottery when $p > 0.5$ and then behave honestly if u_3 is realized. Our results exclude the existence of these types since we find that almost the totality of participants that choose the **Private** lottery when $p > 0.5$ end up cheating (see Figure 4). On the other hand, Kreps’s model has no bite when $p < 0.5$ since the **Public** lottery is always dominated by the **Private** lottery because of monotonicity.

¹⁵This is important because as noticed by Toussaert (2018): “with naturally occurring data, we rarely observe individuals having a preference for a restricted choice set A and yet receiving a larger choice set B.”

This example illustrates how our results can speak to theories assuming preference for flexibility and strengthen the intuition that our participants seem not to be uncertain about their future choice, but rather consistently plan their actions in advance.

6 Conclusions

In this study, we address a key question inspired by Bacon's opening quote: are opportunities that make a thief or rather thieves make their own opportunities? Approaching cheating behavior through the innovative perspective of temptation and commitment *à la* [Gul & Pesendorfer \(2001\)](#), we introduce an original experimental design that permits the identification of individual plans to be honest and to cheat, as well as temptation and costly commitment to honest behaviour.

Our findings speak against Bacon's intuition. We provide no evidence of anticipated temptation, with participants in our experiment not willing to pay a cost to avoid the cheating option. Moreover, we provide evidence of consistent planning both for honest and cheating participants, reinforcing the suggestion that thieves make their own opportunities.

References

- Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153.
- Alós-Ferrer, C., Fehr, E., & Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, *129*(6), 1828–1877.
- Baillon, A., Halevy, Y., & Li, C. (2022). Randomize at your own risk: on the observability of ambiguity aversion. *Econometrica*, *90*(3), 1085–1107.
- Banerjee, R., Baul, T., & Rosenblat, T. (2015). On self selection of the corrupt into the public sector. *Economics Letters*, *127*, 43–46.
- Barfort, S., Harmon, N. A., Hjorth, F., & Olsen, A. L. (2019). Sustaining honesty in public service: The role of selection. *American Economic Journal: Economic Policy*, *11*(4), 96–123.
- Buccioli, A., & Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73–78.
- Dekel, E., Lipman, B. L., & Rustichini, A. (1998). Recent developments in modeling unforeseen contingencies. *European Economic Review*, *42*(3-5), 523–542.
- Dekel, E., Lipman, B. L., & Rustichini, A. (2001). Representing preferences with a unique subjective state space. *Econometrica*, *69*(4), 891–934.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723–733.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547.
- Frederick, S., Loewenstein, G., & O’donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351–401.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin*, *145*(1), 1.
- Gibson, R., Tanner, C., & Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, *103*(1), 532–548.
- Gneezy, U., & Kajackaite, A. (2020). Externalities, stakes, and lying. *Journal of Economic Behavior & Organization*, *178*, 629–643.
- Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, *108*(2), 419–453.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, *1*(1), 114–125.
- Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, *69*(6), 1403–1435.
- Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from india. *American Economic Journal: Economic Policy*, *9*(3), 262–290.
- Heinicke, F., Rosenkranz, S., & Weitzel, U. (2019). The effect of pledges on the distribution of lying behavior: An online experiment. *Journal of Economic Psychology*, *73*, 136–151.

- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, *56*(8), 1645–1655.
- Jacquemet, N., Luchini, S., Rosaz, J., & Shogren, J. F. (2019). Truth telling under oath. *Management Science*, *65*(1), 426–438.
- Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, *102*, 433–444.
- Konrad, K. A., Lohse, T., & Simon, S. A. (2021). Pecunia non olet: on the self-selection into (dis) honest earning opportunities. *Experimental Economics*, *24*(4), 1105–1130.
- Kreps, D. M. (1979). A representation theorem for "preference for flexibility". *Econometrica*, *47*(3), 565–577.
- O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American economic review*, *89*(1), 103–124.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.
- Saccardo, S., & Serra-Garcia, M. (2023). Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, *113*(2), 396–429.
- Schneider, F. H., Brun, F., & Weber, R. A. (2020). Sorting and Wage Premiums in Immoral Work. CESifo Working Paper Series 8456, CESifo.
URL https://ideas.repec.org/p/ces/ceswps/_8456.html
- Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, *22*(s1), S16–S27.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8551.2010.00709.x>
- Soraperra, I., Weisel, O., & Ploner, M. (2019). Is the victim max (planck) or moritz? how victim type and social value orientation affect dishonest behavior. *Journal of Behavioral Decision Making*, *32*(2), 168–178.
- Sutter, M. (2009). Deception through telling the truth?! experimental evidence from individuals and teams. *The Economic Journal*, *119*(534), 47–60.
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length big five and big six personality questionnaires. *Psychological assessment*, *23*(4), 995.
- Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: a lab experiment. *Econometrica*, *86*(3), 859–889.
- Utikal, V., & Fischbacher, U. (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, *85*, 108–111.

A Appendix: Experimental instructions and the Pledge of Honesty

INSTRUCTIONS

Welcome to this experiment.

Please refrain from communicating with other participants for the duration of the experiment and make sure your cell phone is switched off.

Read and follow the instructions carefully; they contain everything you need to know to participate.

If you have a question, please raise your hand. We will come to you and answer your question in private. If the question is relevant for all participants, we will repeat it loudly and answer it.

If you violate these rules, we must exclude you from the experiment and the payout.

You will receive a show-up fee of €7 for your participation.

During the experiment, you can earn an additional amount of money depending on your decisions.

The experiment is divided into two parts.

The **FIRST** part, which will be thoroughly described in the next screens, consists of a series of choices between two options.

The **SECOND** part, which will be described at the end of the first part, consists of a series of different questionnaires/tests.

In the **FIRST** part of the experiment, you will face a series of 11 screens each containing two options: a **PUBLIC LOTTERY** and a **PRIVATE LOTTERY**. You have to choose your preferred lottery between the two.

- The **PRIVATE LOTTERY** will be the same on each of the 11 screens and will give you a probability of 50% to win 15 euros.
- The **PUBLIC LOTTERY**, instead, will be different on each of the 11 screens. It will give you a probability to win the 15 euros that will be decreasing on each of the screens. On the first screen, the **PUBLIC LOTTERY** will give you a probability of 100% to win the 15 euros. On each subsequent screen, the probability will decrease until it reaches 0%.

After you make your choice on each of the 11 screens, the computer will randomly select one of them and you will have the chance to play the lottery you chose on that screen. The experimenter will call your **seat number** and you will go to a separate room where you will play the selected lottery.

The probability of winning the 15 additional euros is not the only difference between the **PUBLIC** and the **PRIVATE** lottery. They are different also in the way in which the lottery will be played if selected by the computer. This difference will be clarified on the next two screens!

Figure 7: Instructions given to the participants - No Charity - Page 1.

How the PUBLIC LOTTERY will be played

The lottery will be played in a different room by each participant separately.

In the room, you will meet an experimenter and the referee.

If on the selected screen you chose the public lottery, the lottery will be played publicly in front of the experimenter. The referee will draw a ball from a bag containing 20 balls.

The balls in the bag will be either Black or White. The proportion of White Balls will be chosen to match the probability of winning of the public lottery.

For instance, if the public lottery has an 80% probability of winning, the bag will be filled with 16 White Balls and 4 Black Balls. Then, after the experimenter has carefully mixed the Balls, the referee will choose one ball from the bag and show it to you and to the experimenter.

If the Ball is White you win 15 euros, if the Ball is Black you win nothing.

How the PRIVATE LOTTERY will be played

The lottery will be played in a different room by each participant separately.

In the room, you will meet an experimenter and the referee.

If on the selected screen you chose the private lottery, the experimenter and the referee will leave the room. On the table, you will find **two** White balls, **one** Black ball, and two bags.

You have to **select two of the three balls and put one ball in each bag**. The third ball must be inserted in the balls' container (a box with multiple White and Black balls that insures that nobody will know the colour of the two balls in the bag).

After completing these two tasks you will call back the experimenter and the referee who will enter the room. The referee will select one of the two bags and extract the ball.

If the Ball is White you win 15 euros, if the Ball is Black you win nothing.

Then, the experimenter and the referee will leave the room, and you will put the two balls back in the balls' container.

Figure 8: Instructions given to the participants - No Charity - Page 2.

INSTRUCTIONS

Welcome to this experiment.

Please refrain from communicating with other participants for the duration of the experiment and make sure your cell phone is switched off.

Read and follow the instructions carefully; they contain everything you need to know to participate.

If you have a question, please raise your hand. We will come to you and answer your question in private. If the question is relevant for all participants, we will repeat it loudly and answer it.

If you violate these rules, we must exclude you from the experiment and the payout.

You will receive a show-up fee of €7 for your participation.

During the experiment, you can earn an additional amount of money depending on your decisions. Furthermore, a charity may receive a donation based on your decisions. The charity is described in the brochure you have received. Please take a minute to examine the brochure.

The experiment is divided into two parts.

The **FIRST** part, which will be thoroughly described in the next screens, consists of a series of choices between two options.

The **SECOND** part, which will be described at the end of the first part, consists of a series of different questionnaires/tests.

In the **FIRST** part of the experiment, you will face a series of 11 screens each containing two options: a **PUBLIC LOTTERY** and a **PRIVATE LOTTERY**. You have to choose your preferred lottery between the two.

- The **PRIVATE LOTTERY** will be the same on each of the 11 screens and will give you a probability of 50% to win 15 euros.
- The **PUBLIC LOTTERY**, instead, will be different on each of the 11 screens. It will give you a probability to win the 15 euros that will be decreasing on each of the screens. On the first screen, the **PUBLIC LOTTERY** will give you a probability of 100% to win the 15 euros. On each subsequent screen, the probability will decrease until it reaches 0%.

After you make your choice on each of the 11 screens, the computer will randomly select one of them and you will have the chance to play the lottery you chose on that screen. The experimenter will call your **seat number** and you will go to a separate room where you will play the selected lottery.

The probability of winning the 15 additional euros is not the only difference between the **PUBLIC** and the **PRIVATE** lottery. They are different also in the way in which the lottery will be played if selected by the computer. This difference will be clarified on the next two screens!

Figure 9: Instructions given to the participants - Charity - Page 1.

How the PUBLIC LOTTERY will be played

The lottery will be played in a different room by each participant separately.

In the room, you will meet an experimenter and the referee.

If on the selected screen you chose the public lottery, the lottery will be played publicly in front of the experimenter. The referee will draw a ball from a bag containing 20 balls.

The balls in the bag will be either Black or White. The proportion of White Balls will be chosen to match the probability of winning of the public lottery.

For instance, if the public lottery has an 80% probability of winning, the bag will be filled with 16 White Balls and 4 Black Balls. Then, after the experimenter has carefully mixed the Balls, the referee will choose one ball from the bag and show it to you and to the experimenter.

If the Ball is White you win 15 euros, if the Ball is Black you win nothing.

Independently from the result, we will donate 2 euros to the charity.

How the PRIVATE LOTTERY will be played

The lottery will be played in a different room by each participant separately.

In the room, you will meet an experimenter and the referee.

If on the selected screen you chose the private lottery, the experimenter and the referee will leave the room. On the table, you will find **two** White balls, **one** Black ball, and two bags.

You have to **select two of the three balls and put one ball in each bag**. The third ball must be inserted into a closed box - the charity box - that you find in the room. This box will not be opened until the end of the experiment.

After completing these two tasks you will call back the experimenter and the referee who will enter the room. The referee will select one of the two bags and extract the ball.

If the Ball is White you win 15 euros, if the Ball is Black you win nothing.

Then, the experimenter and the referee will leave the room, and you will put the two balls back in the balls' container (a box with multiple White and Black balls that ensure that nobody will know the colour of the two balls in the bag).

The role of the third ball: at the end of the experiment, the referee and the experimenter will open the charity box. The number of **White** balls will determine the charity donation. **For each White ball the charity will receive 2 euros, for each Black ball the charity will receive nothing.**

HENCE: the charity will receive the 2 euros donation only if you insert one White ball & one Black ball in the bags.

Figure 10: Instructions given to the participants - Charity - Page 2.

Pledge of Honesty

I promise that I will respect the probabilities described in the decision screens.

Therefore, in the Private Lottery, I will insert one White and one Black ball into the bags.

Signature

Figure 11: The Pledge of Honesty.

B Appendix: Model

In Section 2, we introduced the GS model with the assumption that $v(\mathbf{C}) \geq v(\mathbf{H}_{pr})$, which seems natural in our context. Here, we generalize the model to allow for the temptation to be honest. Therefore, Figure B.1 generalizes Figure 2 allowing for $v(\mathbf{H}_{pr}) > v(\mathbf{C})$.

Recall that an individual is honest when $u(\mathbf{C}) < 0.5$ and he is a cheater when $u(\mathbf{C}) > 0.5$; however, she cheats in the **LG** stage if $u(\mathbf{C}) + v(\mathbf{C}) > u(0.5) + v(\mathbf{H}_{pr})$ and she behaves honestly otherwise. To simplify the exposition, we report the utilities on the menu $U(\{\mathbf{H}_{pr}, \mathbf{C}\})$ for each type in Table B.1. The first four types are those represented in Figure 2, while Types 5 and 6 are novel.

	Utility	Cost	$U(\{\mathbf{H}_{pr}, \mathbf{C}\})$
1	$u(\mathbf{C}) + v(\mathbf{C})$	$v(\mathbf{C})$	$u(\mathbf{C})$
2	$u(\mathbf{C}) + v(\mathbf{C})$	$v(\mathbf{C})$	$u(\mathbf{C})$
3	$u(0.5) + v(\mathbf{H}_{pr})$	$v(\mathbf{C})$	$u(0.5) + v(\mathbf{H}_{pr}) - v(\mathbf{C})$
4	$u(0.5) + v(\mathbf{H}_{pr})$	$v(\mathbf{H}_{pr})$	$u(0.5)$
5	$u(0.5) + v(\mathbf{H}_{pr})$	$v(\mathbf{H}_{pr})$	$u(0.5)$
6	$u(\mathbf{C}) + v(\mathbf{C})$	$v(\mathbf{H}_{pr})$	$u(\mathbf{C}) + v(\mathbf{C}) - v(\mathbf{H}_{pr})$

Table B.1: Utility in the hot state, cost of temptation, and $U(\{\mathbf{H}_{pr}, \mathbf{C}\})$.

Since we allow for the temptation to be honest, we now slightly modify the notation for the types. Type 1 ("cheaters") become "cheaters tempted to cheat" while Type 4 ("honest without temptation") become "honest tempted to be honest". Interestingly, allowing for Types 5 and 6 does not change the predictions in our experiment. Type 5 behave exactly as Type 4 since $U(\{\mathbf{H}_{pr}, \mathbf{C}\}) = u(0.5)$ which is equal to $u(p)$ only for $p = 0.5$. Namely, similarly to Type 4, "cheaters tempted to be honest that given in" choose the **Public** lottery if $p > 0.5$ and **Private** lottery if $p < 0.5$. Type 6, instead, behave as Type 1 but with the only caveat that their switching point is lower than what it would have been without temptation. In fact, notice that $U(\{\mathbf{H}_{pr}, \mathbf{C}\}) = u(\mathbf{C}) + v(\mathbf{C}) - v(\mathbf{H}_{pr})$. Hence, these participants end up cheating if facing the **Private** lottery, however, since they costly anticipate the temptation to be honest, they choose the **Public** lottery more often than Type 1 when $p > 0.5$.

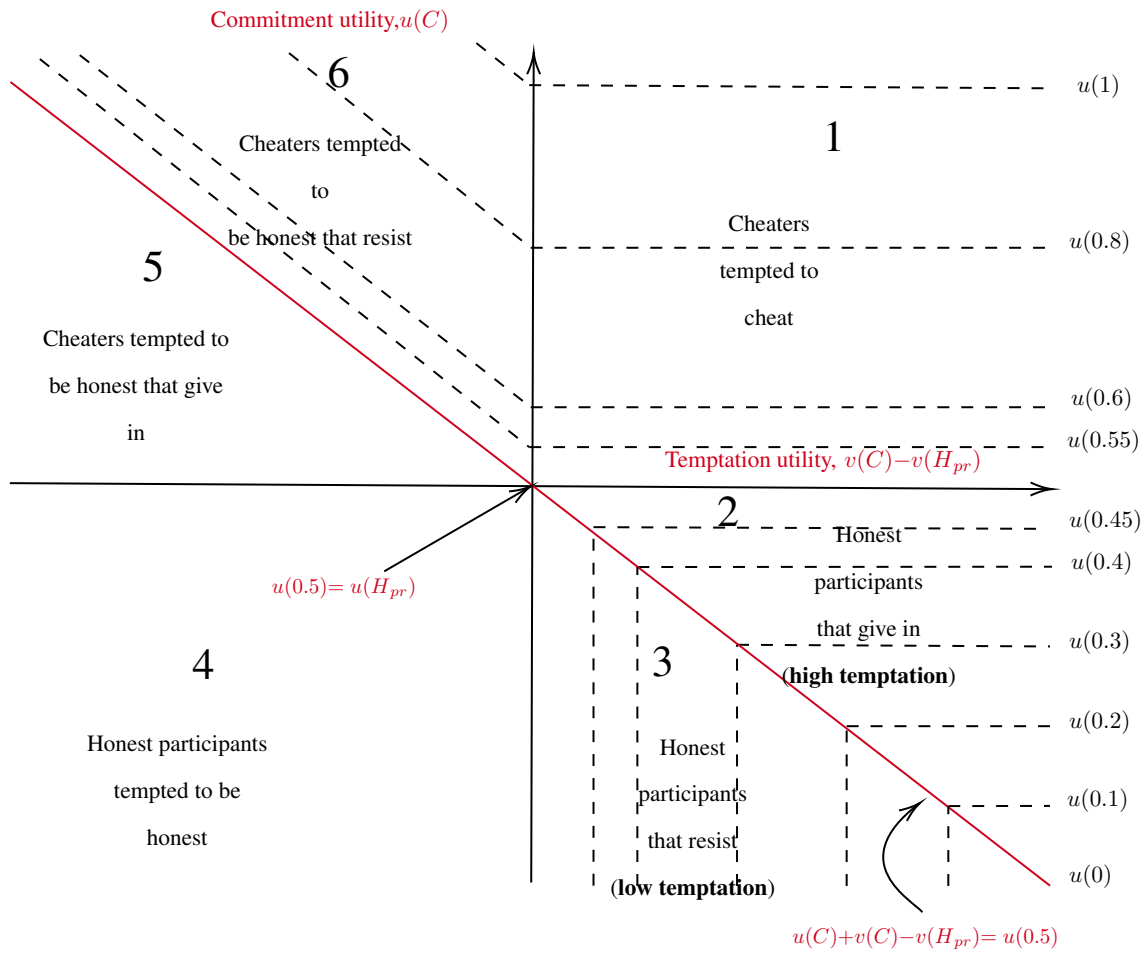


Figure B.1: Types

Notes: The dashed lines represent indifference curves for the different levels of p . The equation of these curves can be simply found solving the equation $U(\{H_{pr}, C\}) = u(p)$ and substituting different expression from Table B.1.

C Appendix: The Private lottery using a coin-flip

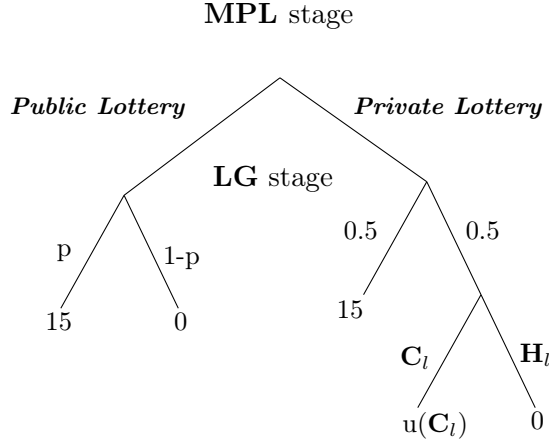


Figure B.2: Extensive form representation of the decision problem using a coin-flip

Figure B.2 represents our experimental set-up as shown in Figure 1 with the only difference that the **Private** lottery is played in standard fashion, reporting the result of a coin flip. In this section, we discuss the theoretical differences between our design and this modified version, henceforth "coin-flip design". First and foremost, as mentioned in Section 5, the coin-flip design requires the assumption that the decision-maker does not lie against her self-interest. This is impossible in our setting where playing honestly implies committing to a 50-50 lottery, while cheating implies a sure win. In Figure B.2, this assumption is embedded in the representation as we assume that when the coin-flip is successful the decision-maker wins 15 euros for sure. The second important difference regards the theoretical nature of the **Private** lottery. For the purpose of this short section, we denote behaving honestly and cheating in the **Private** lottery as \mathbf{C}_l and \mathbf{H}_l respectively given that both behaviours are conditional on losing the coin-flip. Importantly, this is not only a choice of language. There may be good reasons for which $u(\mathbf{C}_l) \neq u(\mathbf{C})$ and $u(\mathbf{H}_l) \neq u(\mathbf{H}_{pr})$. For instance, on the one hand, a decision-maker may find it less morally costly to cheat after having discovered to have lost the coin flip than when having to commit to a sure win; on the other hand, behaving honestly may be more rewarding when it implies a sure loss. In this sense, as anticipated in Section 5, the coin-flip design provides a potentially different measure of cheating. Finally, in the GS model, the **Private** lottery is represented by the following menu $\{(0.5, 15), (0.5, \{\mathbf{C}_l, \mathbf{H}_l\})\}$. Note that, in our design, the utility u is defined on money lotteries and does not require assumptions on the aggregation of utilities. In

this case, the lottery is defined between money and the menu $\{\mathbf{C}_l, \mathbf{H}_l\}$; hence, one has to assume that the same aggregator holds in both types of lotteries. Furthermore, given that the cheating cost arises only in case of a loss, assumptions should exclude probability weighting functions which could potentially introduce concepts of optimism/pessimism and confound the measurement of the cheating cost.

D Appendix: Personality traits

After our main experiment, participants answered a standard Big Six questionnaire with 48 items that yield scores for the following traits (Thalmayer et al. (2011)): Honesty, Agreeableness, Resiliency, Extraversion, Originality, and Conscientiousness.

As shown in Figure B.3, we find no difference in personality traits between participants who switched at $p > 0.5$, $p = 0.5$, or $p < 0.5$. The only exception is Honesty. We find that participants who switch at $p < 0.5$ are significantly more honest (t-test, rank-sum test, $p < 0.01$) than those who switch at $p > 0.5$. We interpret this correlation as a validation of our experimental task.

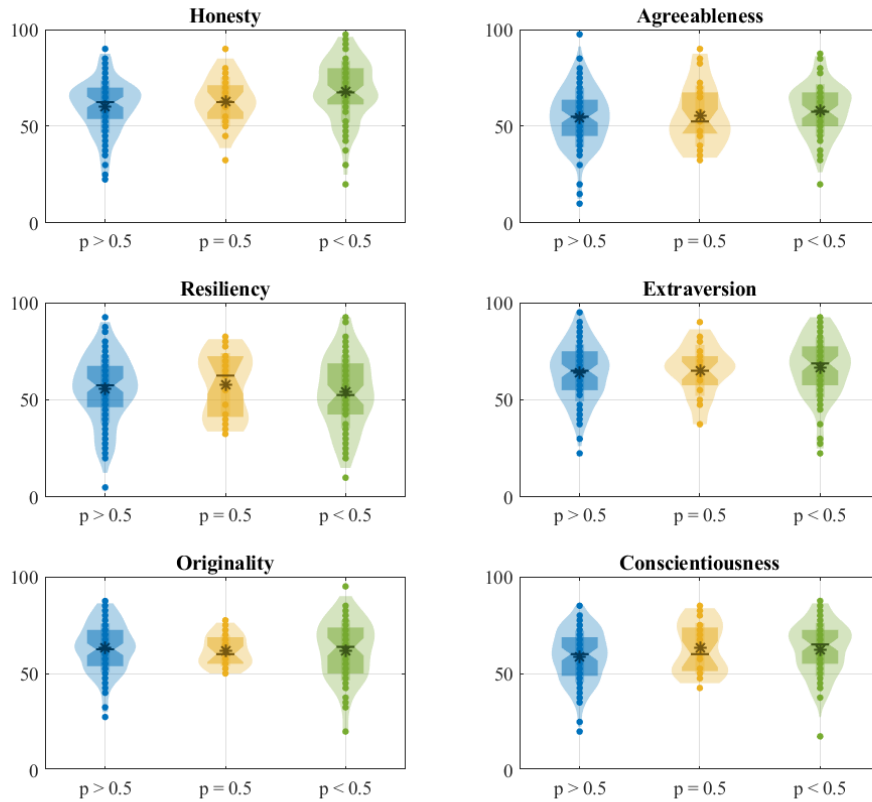


Figure B.3: Big 6 - Personality Traits.

Notes: We report the distributions of normalized Personality traits using violin plots with kernel density plots where * is the mean, - is the median, the box-plot limits are the 25th and 75th percentile.

Discussion Papers of the Research Area Markets and Choice 2023

Research Unit: **Market Behavior**

- Levent Neyse, Frank M. Fossen, Magnus Johannesson, Anna Dreber** SP II 2023-201
Cognitive reflection and 2D:4D: Evidence from a large population sample
- Christian Basteck, Lars Ehlers** SP II 2023-202
On the constrained efficiency of strategy-proof random assignment
- Hande Erkut, Ernesto Reuben** SP II 2023-203
Social networks and organizational helping behavior: Experimental evidence from the helping game
- Tobias König, Lydia Mechtenberg, Dorothea Kübler, Renke Schmacker** SP II 2023-204
Fairness in matching markets: Experimental evidence
- Daniele Caliari, Ivan Soraperra** SP II 2023-205
Planning to cheat. Temptation and self-control

Research Unit: **Economics of Change**

- Kai Barron, Tilman Fries** SP II 2023-301
Narrative persuasion
- Christina Timko, Maja Adena** SP II 2023-302
Transparent app design reduces excessive usage time and increases willingness to pay compared to common behavioral design—a framed field experiment
- Daniele Caliari** SP II 2023-303
Behavioural welfare analysis and revealed preference: Theory and experimental evidence
- Daniele Caliari** SP II 2023-304
Rationality is not consistency
- Steffen Huck, Macartan Humphreys, Sebastian Schweighofer-Kodritsch** SP II 2023-305r
Political salience and regime resilience