**Incentives, ambiguity and gain/loss framing as determinants of cheating: An experiment and outline of a model.**

*manuscript version: 10th Dec 2019.*

Łukasz Markiewicz [1] *

Marcin Malawski [1]

Tadeusz Tyszka [1]

[1] Center of Economic Psychology and Decision Sciences, Kozminski University

* Corresponding author: Łukasz Markiewicz, Center of Economic Psychology and Decision Sciences, Kozminski University, 57/59 Jagiellonska St., 03-301 Warsaw, POLAND, email: lmarkiewicz@kozminski.edu.pl (mobile phone +48 602 222 087)

**Funding**

This work was supported by the National Science Centre, Poland (NCN) under Grant 2015/17/D/HS6/02684.

# Incentives, ambiguity and gain/loss framing as determinants of cheating: An experiment and outline of a model.

Abstract

In an online study we observed that people cheated more frequently when (1) monetary incentives were greater, (2) monetary incentives were framed as losses rather than gains, and (3) situations were more ambiguous. Additionally, we found that (4) reaction times were longer for decisions made in situations involving conflict between material incentives and the norm of being honest, and (5) were longer for more ambiguous stimuli than less ambiguous stimuli, particularly in congruent situations. Confirmation of these hypotheses permitted us to sketch the outline of a model describing decision making in situations involving moral aspects. The proposed model concerns situations where an immoral action goes unpunished and the probability of its detection is either zero or irrelevant to a decision. We assume that in such situations agents do not simply strive to maximize their own material well-being, but, rather, in addition to self-interest, they are also sensitive to the moral aspects of decisions. We believe that decision making processes under the conditions studied involve multiple criteria and tradeoffs similar to those encountered in purely economic decisions, and that these processes obey similar principles.

## INTRODUCTION

The present study focuses upon how people make decisions involving moral issues – in particular, decisions concerning whether to cheat or engage in other actions transgressing moral norms. Becker (1968) proposed a pure economic model, assuming that people decide to transgress if this pays-off. Thus, a decision as to whether to cheat should depend on the potential benefits of cheating (the magnitude of the monetary incentive) and the potential costs (the probability of being caught and the severity of the possible penalty). On the other hand, others depart from Becker's assumption that only monetary utility matters, stressing the role of preserving self-esteem and other moral aspects (for a review, see Jacobsen, Fosgaard, and Pascual-Ezama (2017). We agree with the claim that economic agents are not entirely self-interested and that they do not simply strive to maximize their own material well-being, i.e., that, apart from self-interest, they are sensitive to moral aspects of decision making situations.

Under these assumptions, a search for the determinants of how decisions involving moral aspects are made should focus on two fundamental factors: (1) the benefit accruing to the decision maker (DM) in terms of the monetary payoffs linked to cheating behavior, and; (2) the disutility of a loss in self-esteem.

As far as the first factor is concerned, there is debate as to whether (in line with the common assumption that a rational economic agent should maximize utility) people are more willing to cheat in the presence of greater material incentives or whether people may be insensitive to incentive size and cheat only a little irrespective of incentive size (Mazar, Amir, & Ariely, 2008). Here, empirical results do not provide clear-cut answer. Some previous studies have found no effect of incentives on cheating (Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008; Rahwan, Hauser, Kochanowska, & Fasolo, 2018). However, other studies have found support for the hypothesis that people cheat more when higher incentives are offered (Hilbig & Thielmann, 2017; Kajackaite & Gneezy, 2017; Markiewicz & Czupryna, 2019; Markiewicz & Gawryluk, 2019).

The present experiment considered the effect of incentives on cheating by investigating not only the choices people make, but also the time they take to make them. Mental chronometry has a long tradition in psychological research (Donders, 1969); for a review see Jensen (2006). While previous cheating studies have focused mostly on whether reaction times (RTs) differ for lying and truth telling (Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017), presently we went one step further to research RTs in contexts both where self-serving lies were possible and not possible.

We forwarded two hypotheses:

With respect to decision times, it is reasonable to assume that a person experiencing a conflict between self-interest and socially accepted principles (henceforth referred to as an incongruent situation) should require more time to make a decision than when no such conflict occurs (a congruent situation). Thus, we tested the hypothesis that **decision times should be longer for choices made in incongruent situations compared to those made in congruent situations (Hypothesis 1).**

The second hypothesis stated that, in the absence of any suspicion that cheating behavior will be uncovered, **people should cheat more frequently under higher monetary incentives than under lower monetary incentives (Hypothesis 2)**.

Moreover, given that prospect theory (Kahneman & Tversky, 1979) suggests that "losses should loom larger than gains", it can be conjectured that an incentive framed as a loss will result in greater cheating than the same incentive framed as a gain. We therefore tested the hypothesis that **more cheating should be observed *(ceteris paribus)* under loss framing than under gain framing (Hypothesis 3)**.

However a greater challenge is to investigate people's sensitivity to moral aspects of decisions. This stems from the fact that people's moral values seem to be much more diverse than their material values. As noted by Kajackaite and Gneezy (2017), p. 433:

> Some people may be unwilling to tell a lie, regardless of their benefit from it ("ethical type"), …. People who are not willing to lie could be described in our approach as having an infinite cost of lying. Other people may have a finite positive intrinsic cost of lying. These people will lie when the benefit of lying is higher than the associated cost ("finite positive cost type"); at the extreme are people with a zero cost of lying ("economic type").

Similarly, Hilbig and Thielmann (2017) demonstrated that incentives trigger cheating to different extents among different clusters of people. While "corruptible individuals" easily violate norms when this pays off, "small sinners" do this only up to a certain point and "honest individuals" do not cheat at all, regardless of the magnitude of a monetary incentive.

However, the present study did not consider individual differences. The weight ascribed to moral conduct may depend not only on a DMs' dispositional characteristics but also on situational factors. In particular, one important factor is the level of ambiguity of a decision situation: an ambiguous stimulus can often be used to justify self-serving cheating: doing wrong but feeling morally correct (Pittarello, Leib, Gordon-Hecker, & Shalvi, 2015).

The above considerations led to the hypothesis that **people should cheat more frequently when a situation is more ambiguous (Hypothesis 4)**.

We also considered whether ambiguity affects the time taken to make a decision. We theorized that two types of process should influence decision times: (1) cognitive perceptual processes and; (2) the process of making the decision as to whether or not to cheat. Certainly, the time taken to identify ambiguous stimuli will *per se* be longer than the time taken to identify unambiguous stimuli. However, conflict between self-interest and moral principles (in incongruent situations) may cancel out time differences between the recognition of easy and more demanding ambiguous stimuli, possibly making decision times as to whether to cheat or not cheat similar for ambiguous and unambiguous stimuli. This reasoning led to the hypothesis that **the total decision time for an ambiguous stimulus should be longer than that for unambiguous stimulus in congruent situations, but this should not be true for choices made in incongruent situations (Hypothesis 5).**

The ultimate goal in testing the above hypotheses was to construct a simple but general model explaining cheating frequency as a function of the size of material incentives, loss of self-esteem and ambiguity of decision situations. We present and provide a preliminary discussion of such a model in the final section.

## METHOD

### Participants

Feelings of illusory anonymity foster cheating (Zhong, Bohns, & Gino, 2010). Thus, it is not surprising that cheating has been shown to be greater when the same study is performed online compared to when it is performed in a laboratory (Bereby-Meyer et al, 2018). Interestingly, studies have discussed the possibility that clusters of "cheaters" exist among MTurk samples (Chandler & Paolacci, 2017). In general, MTurkers cheat more than laboratory samples (Gerlach, Teodorescu, & Hertwig, 2019), some authors claiming that dishonesty levels for MTurk samples are among the highest of all the platforms used for crowdsourcing behavioral research (Peer, Brandimarte, Samat, & Acquisti, 2017). Given these findings, we used an MTurk population because this had a distinct advantage given the topic of our research.

Participants were recruited through Amazon Mechanical Turk (AMT) using Turk Prime – now known as CloudResearch (Litman, Robinson, & Abberbock, 2017). To avoid language issues and the influence of cultural aspects we limited participation to US residents (as verified by IP inspection). We also limited participation to reputable respondents, using a minimum 80% approval rating for previous Human Intelligence Tasks (HITs) as an inclusion criterion to exclude potentially problematic participants (Peer, Vosgerau, & Acquisti, 2014).

There was a 94% completion rate. Altogether, 50 people completed the whole task: 27 women and 23 men, aged 23 to 64 years (*M* = 37.52, *SD* = 9.75).

**Design**

The experimental procedure was inspired by the "dots task" (Gino, Norton, & Ariely, 2010) with varying levels of ambiguity (Hochman, Glöckner, Fiedler, & Ayal, 2016). As in previous studies, DMs were asked to state which side of a square (left/right) contained more dots.

The study adopted a full-factorial within-participants design, with cheating as the DV, and three IVs: ambiguity (dots left vs. right: 4/17, 5/16, 6/15, 7/14, 8/13, 9/12, 10/11[1]) x monetary incentive to cheat (no incentive to cheat, 3 cents for an act of cheating, 9 cents for an act of cheating [2]) x frame (loss or gain).

***Independent variable 1: ambiguity***

Each block consisted of 18 trials. Every trial involved presenting a screen with a single square box divided into two parts to a participant. There were 21 dots in the whole box, some

[1] For the most part, each ambiguity level (4/17, 5/16, 6/15, 7/14, 8/13, 9/12) was represented by two stimuli (an original stimulus + its mirror reflection), however, the 10/11 level was represented by six stimuli (three original stimuli + three mirror reflections).
[2] Each temptation level was represented by two blocks (an original block + and a block with a mirror reflection of the original block's payment schema). This symmetry ensured that RTs would not be biased by DMs' right- or left-handedness.

in the right-hand part of the box and some in the left-hand part of the box (see Figure 1 for an example). Boxes differed in ambiguity: from low ambiguity trials with an easily categorizable proportion of 4/17; to high ambiguity trials, where categorization of the dots (10/11) in a box was not obvious.

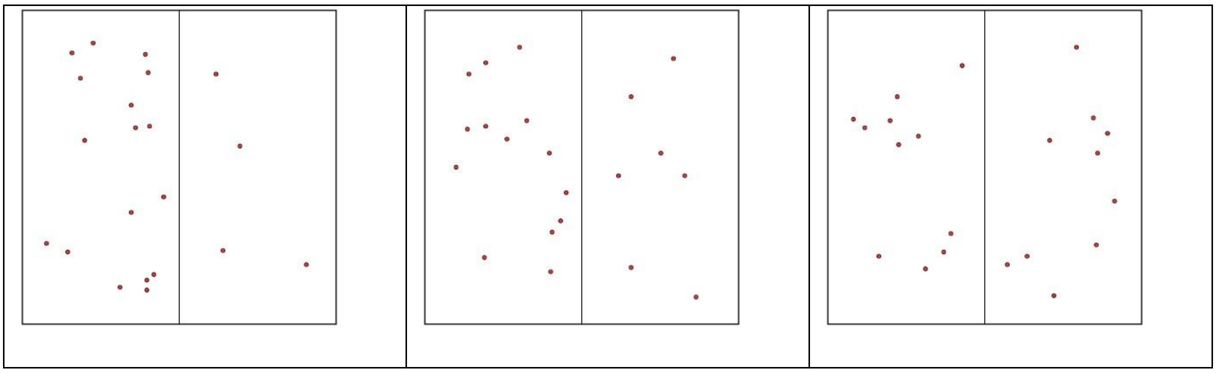

Figure 1. An example of three levels of stimulus ambiguity presented to respondents: left panel, low ambiguity (17/4); middle panel, moderate ambiguity (14/7); right panel, high ambiguity (11/10). All the boxes presented here have the majority of dots located on the left-hand side.

In each trial, a participant’s task was to indicate which side of the box contained more dots.

***Independent variable 2: monetary incentive***

Participants were informed that they would be paid for their responses. They were also informed that the amounts associated with LEFT and RIGHT responses might be unequal (thus, the payoff for an answer indicating "more on the left" may have been higher or lower than for one indicating "more on the right").

Participants could encounter three types of different disproportions of payments between the LEFT and RIGHT categorizations:

- no monetary incentive to cheat (with the same payment of 1¢ for both answers)

- a low monetary incentive to cheat (with one of the answers four times more profitable [4¢] than the other [1¢] – thus, the monetary incentive to cheat was 3¢),
- a high monetary incentive to cheat (with one of the answers ten times more profitable [10¢] than the other [1¢]– thus, the monetary incentive to cheat was 9¢).

So, in trials where there was a stimulus with more dots on one side of the box but in which participants were actually paid more (4¢ or 10¢ vs. 1¢) for the answer that there were more dots on the other side, the task presented a conflict between giving an accurate answer and maximizing profit. Using implicit-association test (IAT) terminology (Greenwald, Nosek, & Banaji, 2003), such rounds can be called "incongruent" (or incompatible) in contrast to "congruent" (compatible) rounds with more profitable honest answers.

Although Rosenbaum, Billinger, and Stieglitz (2014) considered the pros (fewer problems with maintaining one's self-image when cheating) and cons (small stake sizes may be insufficiently high to deter intrinsically dishonest behavior) of using small stakes in cheating studies, in a paper with the suggestive title "Lie for a Dime", Chandler and Paolacci (2017) demonstrated that internet populations are eager to lie to earn even small incentives (people misreported characteristics relevant to meeting explicitly stated eligibility criteria for a study just to earn a "dime"). Moreover, in their original study, Gino et al. (2010) used even smaller incentives (0.5¢ vs. 5¢), while Hochman et al. (2016) used exactly the same incentives as our study (1¢ vs. 10¢). Thus, we believed that even small stakes such as those used in our study would be high enough to observe cheating and cheating differences.

### *Independent variable 3: gain vs. loss framing*

The whole task was repeated twice in random order; once for a GAIN task and once for a LOSS task. Participants started the GAIN task with no initial capital and were informed that for each RIGHT or LEFT answer they would receive a payment. In the LOSS task they were endowed with initial capital of $6.50 from which a certain amount was deducted after each RIGHT or LEFT answer. The final value of a participant's portfolio (as indicated by a dynamic progress bar, updated after each answer: from left [$0] to right [$6.50] in the gain condition and from right [$6.50] to left [$0] in the loss condition) showed a participant's payment for this part of the study. In both conditions, participants were assured that they would obtain their payoffs even if they were not correct in their judgments.

In the GAIN condition, participants started with initial capital of 0$, obtaining $3.24 if they answered honestly (i.e., always correctly), or obtaining $5.40 if they (always) answered dishonestly. In the LOSS condition, participants started with initial capital of $6.50, obtaining $3.26 if they (always) answered honestly, or obtaining $5.42 if they (always) answered dishonestly. This symmetry ensured that honest participants would end-up with the same final balance in each framing condition, and the same was true for dishonest participants.

**Procedure**

The study protocol was approved by Kozminski University's Ethical Committee, and the study was advertised on AMT as a "Recognition and memory study". Before participation, participants were informed that we would present some visual stimuli to them and ask them some related questions. Payments were advertised "depending on participants' luck and performance". The procedure was scripted in Inquisit 4.0.10 (2015); see supplementary online materials for the script: osf.io/6b3jx) and was divided into:

- an initial practice "block 0" (not associated with any payoff: after each answer, participants were simply told the time their answer took in milliseconds and given no feedback relating to the correctness of their answer).

- followed by six blocks (presented in random order) of the GAIN (LOSS) task
- followed by six blocks (presented in random order) of the LOSS (GAIN) task.

A nested structure was adopted, tasks for the three monetary incentive levels being nested within each type of framing[3], and the three ambiguity levels being nested within each incentive level. The procedure is illustrated in Table 1. Blocks 1, 4, 7 and 10 were "no incentive to cheat" blocks, Blocks 2,5,8,11 involved a low incentive to cheat (3¢), and Blocks 3,6,9,12, a high incentive to cheat (9¢). Each of these blocks consisted of 18 randomly ordered trials involving ambiguous stimuli, with ambiguity varying within each block – in each block, half the trials had more dots on the left and half had more dots on the right (mirror reflections).

Participants were instructed to pay attention to the current values of payoffs, which were specific for each block. The order of the LOSS and GAIN tasks was randomized, as was the order of the six blocks within these tasks.

On each of the 216 test trials[4] (12 test blocks x 18 trials) participants saw a screen for 1000 ms with a square box divided into two parts and 21 dots distributed between the two parts. They were asked to indicate where the majority of dots were located (on the left or right), across both "incongruent" trials (dishonest, or false, answer more profitable) and "congruent" trials (honest, or correct, answer more profitable).

[3]To control right hand dominance for RT measurements, each type of block was presented twice, with a reversed payment schema for one presentation: once with the more profitable option under the LEFT key and once, as a separate block, with the more profitable option under the RIGHT key. For example, Blocks 3 and 6 both concerned the "high temptation condition" for the gain frame, these blocks later being combined for analysis.
[4] This was lower than the 300 test trials in the "perceptual task" of Gino et al. (2010), but more the 126 trials in the Hochman et al. (2016) study.

Table 1. The event sequence

| | Block no. | Left Key | Right Key | Stimulus |
|---|---|---|---|---|
| **PRACTICE** | 0 | More on LEFT | More on RIGHT | 18 ambiguous stimuli in random order |
| **GAIN** | 1 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| *(Blocks* | | collect 1 ¢ | collect 1 ¢ | |
| *1 to 6* | 2 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| *random* | | collect 1 ¢ | collect 4 ¢ | |
| *order)* | 3 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| | | collect 1 ¢ | collect 10 ¢ | |
| | 4 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| | | collect 1 ¢ | collect 1 ¢ | |
| | 5 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| | | collect 4 ¢ | collect 1 ¢ | |
| | 6 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| | | collect 10 ¢ | collect 1 ¢ | |
| **LOSS** | 7 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| *(Blocks* | | lose 1 ¢ | lose 1 ¢ | |
| *7 to 12* | 8 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| *random* | | lose 1 ¢ | lose 4 ¢ | |
| *order)* | 9 | More on LEFT | More on RIGHT | The same 18 ambiguous stimuli in random order |
| | | lose 1 ¢ | lose 10 ¢ | |

| | | | |
|---|---|---|---|
| 10 | More on LEFT<br>lose 1 ¢ | More on RIGHT<br>lose 1 ¢ | The same 18 ambiguous stimuli in random order |
| 11 | More on LEFT<br>lose 4 ¢ | More on RIGHT<br>lose 1 ¢ | The same 18 ambiguous stimuli in random order |
| 12 | More on LEFT<br>lose 10 ¢ | More on RIGHT<br>lose 1 ¢ | The same 18 ambiguous stimuli in random order |

The limited time exposure assured that participants did not have time to count the dots and that they therefore only had a general impression of the correct answer. With timing starting at the first millisecond of exposure, participants had to identify which side of the box (right or left) contained more dots by pressing either a letter "E" indicating "more dots on the left", or a letter "I" indicating "more dots on the right" on a computer keyboard. Participants were instructed to go as fast as they could while trying "to make as few mistakes as possible, but giving speed priority over accuracy". It should be noted, however, that Shalvi, Eldar, and Bereby-Meyer (2012) have shown that time pressure can be seen as a mitigating factor for cheating, and the present approach could possibly be seen as diminishing participant blame for producing cheating answers – as suggested by Rosenbaum et al. (2014): respondents could claim that they answered incorrectly due to the time-limited exposure of stimuli, but that they had no intention of cheating.

After each trial, participants received feedback about their reaction time (RT) and earnings for the trial, and their total cumulative earnings were reflected by updating the progress bar.

Once participants had completed the experimental procedure (which took 15.26 minutes on average, *SD* = 3.44 minutes) the script directed them to input demographic information, complete one open-ended feedback question and another questionnaire survey which was not relevant to this study. Later, participants were paid according to the sum accumulated in both (gain and loss frame) “perception tasks”.

## RESULTS

The total value of payments expected for fully honest and perfectly observant participants was \$3.24 x 50 (participants) x 2 (conditions) = \$324.00. The actual payment for the 50 participants amounted to \$458.59. This shows that participants took an extra \$134.59 (41.54%) from our research budget. This observation alone clearly shows that cheating was quite common.

The payoff frequency function (see Figure 2) confirms the above conclusion and, moreover, demonstrates that the two most numerous groups of participants were: (1) those cheating only a little or not at all, and; (2) those cheating almost always, i.e., those focused only on maximizing their own profit regardless of the ethical barriers they had to break. The existence of such clusters of DMs, who take the same action regardless of circumstances, is likely to impede the study of determinants of cheating. However, despite this, all the independent variables had a significant impact on all of our proxies of the propensity to cheat.

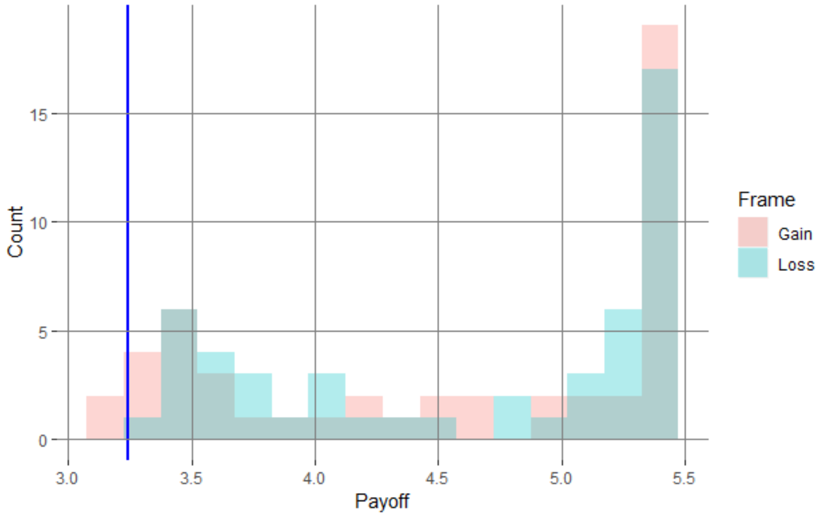

Figure 2. Counts of final payoffs for the total sample. The blue and pink bars represent the distributions for loss and gain framing respectively. The blue line represents the total payoff for an honest and cognitively capable participant.

Since some authors (e.g., Jaeger (2008) have doubted the feasibility of using ANOVA for analyzing data produced by studies such as the current one, data were analyzed using a mixed models approach (Judd, Westfall, & Kenny, 2012) within the R environment (R Core Team, 2017). Specifically, raw trial-wise choices were analyzed in the form of a generalized linear mixed model permitting the analysis of repeated measures in regression models using the glmmTMB package of Brooks et al. (2017). Because a priori power analysis for a generalized linear mixed model is not currently available, as a rough guide, the required sample size for a repeated measures ANOVA was calculated instead using GPower (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). This showed that a sample size of $N = 38$ ($N = 50$) was required to detect a medium effect ($f = 0.25$) for 18 trials (12 trials with non-equal LEFT-RIGHT payments), with $\alpha = .05$ and 95% power. As mixed models control for variance within particular subjects, the required sample size may in fact be smaller when analyzing data using a mixed model instead of using ANOVA. Thus, our sample can be considered sufficient to detect medium to large effects.

Binary decisions were used as DVs in three models, and the logit of this as a binomial link function:

Model 1 involved the simplest and most straightforward DV which did not require dropping any trials from the analysis: **$DV_1$ was a factor variable coded 0-1, taking a value of 1 when a DM gave a correct answer in experimental blocks.**

However, using correct reports may not be the optimal approach for studying cheating since people could potentially have erred in both directions: toward self-serving lies and toward self-threatening mistakes. Thus, Model 1 was re-formulated using a binary variable indicating whether DMs chose the more profitable option as a DV. Gino et al. (2010) and Mazar and Zhong (2010) also used this approach, using the number of times participants selected a high paying response as a proxy for dishonest behavior. For this model (Model 2), in all trials involving the same payment (positive or negative) for left-hand and right-hand classifications (1 cent vs. 1 cent), there was no possibility of knowing which choices were self-serving choices, therefore data for these trials were omitted from analysis: **$DV_2$ was a factor variable coded 0-1, taking a value of 1 when a DM gave an answer associated with the higher payoff.** Note that this variable excluded trials with equal left – right payoffs (33% of all trials) and therefore the three levels of the monetary incentive variable were reduced to two levels.

Hochman et al. (2016) suggested categorizing respondents' answers into four categories as follows:

| | Conflict between positive self-esteem and the desire to increase personal gain | No conflict |
|---|---|---|
| Low paying option | Correct rejections | Detrimental errors |
| High paying option | Beneficial errors | Correct hits |

- correct hits – where a person chooses an accurate response that is also high paying (no conflict between positive self-esteem and the desire to increase personal gain);

- correct rejections – where a person chooses an accurate response that is low paying (conflict between positive self-esteem and the desire to increase personal gain);
- beneficial errors – where a person chooses an inaccurate response that is high paying;
- detrimental errors – where a person chooses an inaccurate response that is low paying.

After filtering out all trials with equal payments on each side (resulting in 144 trials remaining out of the 216 total trials), respondents' answers were classified into these four categories. The results showed that correct hits were the most frequent type of answer (47.78%), followed by beneficial errors (32.35%) and correct rejections (17.65%). The number of detrimental errors was small, on average amounting to only 2.22% of all answers given by respondents. In the final model (Model 3) we investigated whether respondents committed beneficial errors in incongruent trials where a more profitable answer required violation of the moral norm: **$DV_3$ was a factor variable coded 0-1, taking a value of 1 when a DM made a beneficial error (giving an answer that was incorrect and associated with the higher payoff) and taking a value of 0 when a DM gave a correct rejection** (note that this variable excluded trials with equal left – right payoffs on congruent trials that did not present any moral conflict for a DM: trials only permitting the possibility of correct hits and detrimental errors).

In all models the IVs were: task frame (factor), monetary incentive measured in cents across left/right payments (0¢, 3¢, 9¢), ambiguity (the number of dots in the part of the square with the lower number of dots). Each time, participant identity (ID) was also included as a random effect (for the intercept). The models can therefore be represented by the following equation specifying the probability of a binary response coded as 1 and a score function *s*.

$$p(lied) = \frac{1}{1 + e^{-s}}$$

where

$$s = \alpha_i + \sum_{j=1}^{3} \beta_j \times X_{ji}$$

In this formula $\alpha_i$ is an individual intercept (modeled as a random effect), $X_{1i}$ is a factor variable representing frame, $X_{2i}$ represents the monetary incentive in cents, and $X_{3i}$ represents ambiguity.

Table 2 presents the results for a logistic mixed effects model where the probability that DV = 1 is predicted by a set of independent variables and a random intercept for each participant. The odds ratios (ORs) represent the odds that DV = 1 occurred given a particular circumstance compared to the odds of DV = 1 occurring in the absence of that circumstance. Thus, odds ratios above 1 mean that a certain circumstance was associated with higher odds of DV = 1.

Table 2. Results for a generalized linear mixed effects model: DV = a binary variable

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| **DV:** | **Correct choice** | | | **More profitable choice** | | | **Beneficial error** *(incorrect & more profitable choice)* | | |
| **Base:** | **All trials** | | | **Trials with unequal left/right payments (both incongruent and congruent)** | | | **Incongruent trials only** | | |
| *Predictors (IVs:)* | *Odds ratio* | *CI* | *p* | *Odds ratio* | *CI* | *p* | *Odds ratio* | *CI* | *P* |
| (Intercept) | 29.66 | 22.92 – 38.37 | **<.001** | 3.33 | 1.66 – 6.66 | **.001** | 0.06 | 0.02 – 0.17 | **<.001** |
| Frame (Gain) | 1.10 | 1.01 – 1.20 | **.036** | 0.81 | 0.71 – 0.93 | **.003** | 0.65 | 0.52 – 0.80 | **<.001** |
| Monetary incentive in cents | 0.90 | 0.89 – 0.91 | **<.001** | 1.05 | 1.02 – 1.07 | **<.001** | 1.10 | 1.06 – 1.14 | **<.001** |
| Ambiguity | 0.78 | 0.76 – 0.79 | **<.001** | 1.15 | 1.12 – 1.19 | **<.001** | 1.75 | 1.65 – 1.86 | **<.001** |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 3.29 | | | 3.29 | | | 3.29 | | |
| $\tau_{00}$ | 0.33 $_{subject}$ | | | 4.77 $_{subject}$ | | | 10.56 $_{subject}$ | | |
| ICC | 0.09 $_{subject}$ | | | 0.59 $_{subject}$ | | | 0.76 $_{subject}$ | | |
| Observations | 10800 | | | 7200 | | | 3600 | | |
| Marginal $R^2$/ Conditional $R^2$ | 0.111 / 0.191 | | | 0.015 / 0.598 | | | 0.104 / 0.787 | | |

In Model 1 (a model involving all trials, the left-hand column in Table 2), a higher monetary incentive and ambiguity decreased the odds of reporting a correct answer (both ORs < 1), while gain framing increased the odds (OR > 1). Similarly in Model 2 (involving only trials  with unequal left/right payments) a higher monetary incentive and ambiguity increased the odds of choosing a more profitable answer (both ORs > 1), while gain framing decreased the odds (OR < 1). An identical pattern also occurred in Model 3 (involving only incongruent trials with unequal left/right payments where choosing the more profitable answer required dishonesty): a higher monetary incentive and ambiguity increased the odds of committing a beneficial error (both ORs > 1) while gain framing decreased the odds (OR < 1).

Thus, in Model 3, for each cent increase in monetary incentive and with each dot moved from the majority to the minority side of the square thus increasing ambiguity, the odds of committing a beneficial error increased by factors of 1.10 and 1.75 respectively. This indicates that a DM seeing an extra dot on the minority side was 1.75 times more likely to commit a beneficial error than a DM not seeing an extra dot, this supporting H4. Similarly a DM exposed to an extra cent of monetary incentive was 1.10 times more likely to commit a beneficial error than a DM not so exposed, this supporting H2. As an odds ratio less than 1 implies a negative relationship, exposing a DM to gain framing corresponded with lower odds of committing a beneficial error. Or, in other words, being exposed to loss framing, increased the odds of a DM committing a beneficial error by 1.54 times (1/0.65), this supporting H3.

Note that the virtual absence of detrimental errors made all of the above analyses very similar. In fact, if people made no detrimental errors at all (DE = 0 for every frame / ambiguity / non-zero incentive configuration), then any of the three dependent variable measures of cheating would uniquely determine the other two by the linear equations $C = M - BE$ and $B = \frac{M}{2} + BE$, where M is the number of all trials in the given configuration (12 for high ambiguity 10 / 11 and 4 for all other levels of ambiguity), and C, B and BE denote the numbers of correct reports, self-serving reports and beneficial errors respectively, However, the respondents did commit some “genuine” errors as evidenced by the presence of some incorrect reports in the no incentive condition (16.42% of all answers in the no-incentive condition; 2.75% under no ambiguity – up to 6 /15 dots – as opposed to 36% under high ambiguity), and therefore separate analyses are presented for each of the three proxies of propensity to cheat.

**Analysis of reaction times (RTs)**

Self-maintenance theory suggests that decision times should be longer for choices made in incongruent trials compared to those made in congruent trials (Hypothesis 1). An increase in RT should be observed when a person experiences a conflict between positive self-esteem and the desire to increase personal gain (incongruent trials) as compared to trials where more profitable responses do not impede positive self-esteem (congruent trials).

To investigate RTs, a repeated measures regression (mixed model) approach was again adopted using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2017). The distribution of RT data among trials with unequal left/right payments (both incongruent and congruent) was investigated. Only 12 out of 7200 answers (ca. .002%) took longer than 3000ms. Just under one third of answers (31.47%) had RTs below 300 ms, indicating responses given without stimulus recognition (Greenwald et al., 2003) – or pre-planned cheating. Thus, the data were filtered to include only RTs above this conventional 300 ms threshold (Greenwald et al., 2003). RTs were transformed logarithmically to remove skewness in the data as advised by Greenwald et al. (2003). Table 3 presents the model concerning log(RT) in relation to trial characteristics.

Table 3: Linear mixed effects model results: DV = a log RT.

Base: only trials with RT > 300 ms

| | **Model 4** | | | **Model 5** | | | **Model 6** | | |
|---|---|---|---|---|---|---|---|---|---|
| **DV:** | **Log(RT)** | | | **Log(RT)** | | | **Log(RT)** | | |
| **Base:** | **Trials with unequal left/right payment (both incongruent and congruent)** | | | **Incongruent trials only with unequal left/right payment** | | | **Incongruent trials only with unequal left/right payment - correct answers only** | | |
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 6.11 | 6.05 – 6.18 | **<.001** | 6.10 | 6.03 – 6.18 | **<.001** | 6.05 | 5.97 – 6.14 | **<.001** |
| Monetary incentive in cents | 0.00 | -0.00 – 0.00 | .500 | 0.00 | 0.00 – 0.01 | .074 | 0.01 | 0.00 – 0.01 | **<.001** |
| Ambiguity | 0.02 | 0.01 – 0.02 | **<.001** | 0.02 | 0.01 – 0.02 | **<.001** | 0.04 | 0.03 – 0.04 | **<.001** |
| Congruent trial (factor) | -0.15 | -0.21 – -0.09 | **<.001** | | | | | | |
| Ambiguity* Congruent trial (factor) | 0.01 | 0.00 – 0.02 | **.004** | | | | | | |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.09 | | | 0.09 | | | 0.06 | | |
| $\tau_{00}$ | 0.03 subject | | | 0.03 subject | | | 0.04 subject | | |
| ICC | 0.27 subject | | | 0.27 subject | | | 0.38 subject | | |
| Observations | 4928 | | | 2445 | | | 1263 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.030 / 0.293 | | | 0.011 / 0.276 | | | 0.058 / 0.414 | | |

For Model 4, RTs for congruent trials were lower than for incongruent trials, supporting H1. For greater reader convenience, Figure 3 provides an additional illustration of this relationship using non-logarithmically transformed reaction times (mean RTs across particular categories with RTs averaged for each DM).

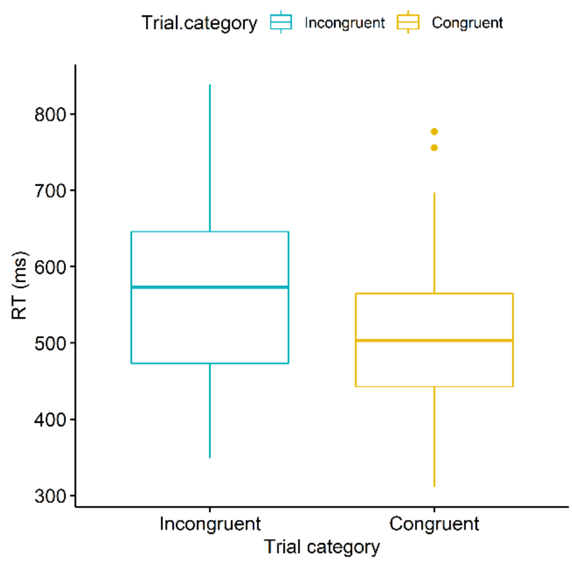


Figure 3. Boxplot for non-logarithmically transformed RTs for Congruent (no conflict) and Incongruent (moral conflict) trials. Base: only trials with RTs > 300 ms

The Model 4 estimate for ambiguity was positive and significantly different from zero, showing that overall RTs increased with the experiencing of ambiguity. We also hypothesized that decision times for ambiguous stimuli should be longer than those for unambiguous stimuli in congruent situations, but not for choices made in incongruent situations (Hypothesis 5). The significant ambiguity x congruent trial interaction term supported this hypothesis. The nature of this interaction is clarified in Figure 4, which shows that while ambiguity did influenced RTs in both types of trial, it had a stronger influence in congruent trials than in incongruent trials.


Figure 4. A depiction of the Model 4 interaction term

As a post hoc analysis, in Model 5 (see Table 3) we also tested whether magnitude of monetary incentive influenced RTs in incongruent trials (where there was a conflict between morality and personal gain). We suspected that the time needed to solve the conflict between maintaining positive self-esteem and the desire to increase personal gain should be longer for highly conflicting incongruent trials (those where a large difference between left/right payments gave a high monetary incentive to cheat) than for moderately conflicting incongruent trials (those where a low difference between left right/payments gave a low monetary incentive to cheat). And indeed trials with higher monetary incentives had longer RTs than those with lower monetary incentives. However, this difference was marginally nonsignificant ($p$ = .089), although it became statistically significant after restricting the sample to correct answers only (Model 6).

So, DMs were sensitive to monetary incentives. Higher monetary incentives led not only to a higher probability of committing beneficial errors as previously shown, but also to longer reaction times in incongruent trials. Thus, as predicted by Rahwan et al. (2018), higher stakes were more “psychologically taxing”.

**DISCUSSION**

The present results enabled us to confirm a number of hypotheses concerning human behavior in situations inviting the breach of a moral norm (cheating) to achieve a material benefit. We observed that people cheated more frequently: (1) **when the monetary incentives were greater (H2)**; (2) **when the monetary incentives were framed as losses rather than as gains (H3)**, and; (3) **when the situation was more ambiguous (H4)**. Additionally, we tested two hypotheses relating to reaction times in situations where people had to make a quick decision as to whether or not to cheat. The reaction times turned out to be **longer for decisions made in situations involving conflict between a material incentive and the norm of honesty (incongruent situations; H1)** and **to be longer for more ambiguous than for less ambiguous stimuli, particularly in congruent situations (H5).**

Three of the hypotheses which found support in our experiment have been tested in earlier studies. Two of them (H3 and H4) have also enjoyed unambiguous support in this previous research. In particular, the (expected) observation of more cheating for payoffs framed as losses rather than gains (H3) is in agreement with recent studies. In a multi-round game (25 dice throws), Schindler and Pfattheicher (2017) asked people to count their number of winning 4s. People cheated more frequently when faced with a loss frame compared to a gain frame. Similarly, using a “matrix task”, Grolleau, Kocher, and Sutan (2016) modeled ex post payments as a gain frame and advance payments as a loss frame, and showed more cheating in the latter. These studies, however, could only identify dishonest behavior at an aggregate level by comparing empirical distributions to expected outcomes (being unable to gather knowledge of exactly who lied and how much). But the recent studies of Markiewicz and Czupryna (2019), Markiewicz and Gawryluk (2019), and Leib, Pittarello, Gordon-Hecker, Shalvi, and Roskes (2019) have detected whether participants cheat at an individual level, permitting the measurement of asymmetries in gain/loss dishonesty at this level. Our study contributes to this line of research.

We also confirmed the hypothesis that cheating is more frequent in more ambiguous situations (H4). Two different phenomena might possibly explain this observation. First, some incorrect answers given by our participants in ambiguous incongruent situations did not necessarily seem to constitute “cheating”, but, rather, might have been genuine errors, as suggested by a considerable number of similarly incorrect answers where ambiguity was high but no monetary incentives were present. Second, however, and more importantly, we also attribute this result to the fact that an ambiguous stimulus can be used to justify self-serving cheating. In fact, as noted by Pittarello et al. (2015), the process underlying people’s self-serving mistakes is not necessarily conscious: people may engage in self-deception unconsciously. In our study, we observed significant differences in reaction times for answers given under conditions of temptation (incongruent trials) and no temptation (congruent trials). Moreover, the same held for RTs where people decided to give an honest answer.

Although significant, these differences concerned reaction times of milliseconds, these times being way below the threshold that would allow us to say that decisions were thoughtful and well considered. This might support the idea that people can engage in self-deception unconsciously.

The main focus of our study was on whether people cheat more frequently under higher than under lower monetary incentives. The results of previous studies as to whether willingness to cheat depends on the size of a material incentive were inconclusive, and this is why we devoted more attention to this matter in the current study. Apart from testing the effects of a monetary incentive’s magnitude on frequency of cheating, we also recorded decision making times in various choice situations.

We obtained support for the hypothesis that people cheat more willingly under higher than under lower monetary incentives. This finding contrasts with those of Fischbacher and Föllmi-Heusi (2013) and Mazar et al. (2008). Also, meta-analyses of studies focusing on the roll a die paradigm have shown no relationship between incentive size and willingness to lie (Abeler, Nosenzo, & Raymond, 2019; Gerlach et al., 2019)[5]. On the other hand, the current findings accord with the behavior of subsets of participants in Hilbig and Thielmann (2017) study and Kajackaite and Gneezy's (2017) mind game study.

Moreover, support for the importance of monetary incentives is also offered by our analysis of reaction times. First, decision times were generally longer in incongruent than congruent trials (Model 4). Second, and more interestingly, within incongruent trials the size of monetary incentives had a marginally nonsignificant influence on reaction times in general but had a highly significant influence on times taken to give correct, and thus less profitable, responses (Model 6). This reflects the common sense notion that it is simply more difficult to resist a stronger temptation than a weaker one.

It is worth stressing that, in contrast to most of the earlier studies focusing on the effects of incentive magnitude on cheating (e.g., (Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008), our results were obtained using a within-subjects design, thereby comparing the decisions and cheating frequencies of the same person in situations differing with respect to material incentives, ambiguity and framing. This probably helps explain why we found an effect where others did not (Charness, Gneezy, & Kuhn, 2012). Additionally, Kajackaite and Gneezy (2017) suggested that previous findings showing no effect of incentive on cheating could have occurred not because the existence of intrinsic cheating costs prevents DMs from cheating more, but because of a DM's suspicion that taking more would expose them and their cheating behavior.

[5] Although the meta-analytic study of Gerlach et al. (2019) also showed that greater potential reward sizes are linked to higher cheating in sender-receiver games.

Moreover, our study's support for hypotheses H2, H3 and H4 is stronger than can be inferred from observing mere statistical significance. Our data were collected from a general population, including people ready to "lie for a dime" irrespective of the circumstances. As can be easily seen from Figure 2, among our participants there was a considerably sized cluster of people (close to 30% of the sample), who (to state it cautiously) could be considered to be *homo oeconomicus* type persons: those who simply chose the response with a higher payoff whenever possible. Here, 19 out of 50 people (38%) gave 3 or less self-harming answers in 36 incongruent gain trials, the same number of people behaving similarly in 36 loss trials. Around 14 DMs (28%) adopted this strategy consistently across both the loss and gain tasks, thus preventing detection of any reaction to ambiguity, frame or size of material incentive (a similar conclusion can be drawn from the fact that 31.47% of responses on trials with unequal left/right payments [both incongruent and congruent] had RTs below 300 ms). These participants were taken into account in our analyses: had we excluded them, we would have obtained far lower *p* values for people showing a reaction to our independent variables.

Confirmation of our five hypotheses allows us to sketch an outline of a model of decision making in situations involving moral aspects. Our model's design concerns situations where an immoral action is not punished and the probability of its detection is either zero or irrelevant to a decision. In view of both our and many earlier findings, we assume that agents do not simply strive to maximize their own material well-being, but that they are also sensitive to the moral aspects of their decisions. We also assume that moral attributes can be incorporated into the expected utility model accepted in modern decision theory.

Thus, when a DM places a positive value on the moral aspects of behavior (e.g., on the well-being of others), their utility function should include both the utility of a material gain and the disutility associated with the possible breach of a moral norm. Assuming that this disutility depends on the size of the breach expressed as a function of the material incentive (as is true in the case of our study – cheating for a cent not being equivalent to cheating for a dime), total utility $U$ can be expressed as:

$$U(x) = (1-\alpha)\, u_s(x) - \alpha\, u_m(x) \qquad (1)$$

where the variable $x$ denotes magnitude of temptation and represents the material gain from committing an immoral action (e.g., cheating), $u_s(x)$ is the benefit accruing to a DM from outcome $x$, $u_m(x)$ is the disutility of a loss in self-esteem caused by committing the action to obtain the material reward $x$, and **α** expresses the DM's concern for moral aspects.

This model is able describe a wide variety of behaviors and permits the formation and testing of empirical hypotheses concerning factors determining cheating behavior. Assuming, for instance, a classical increasing and convex utility function $u_s$ and an increasing linear disutility function $u_m$ we obtain the curves shown in Figure 5.

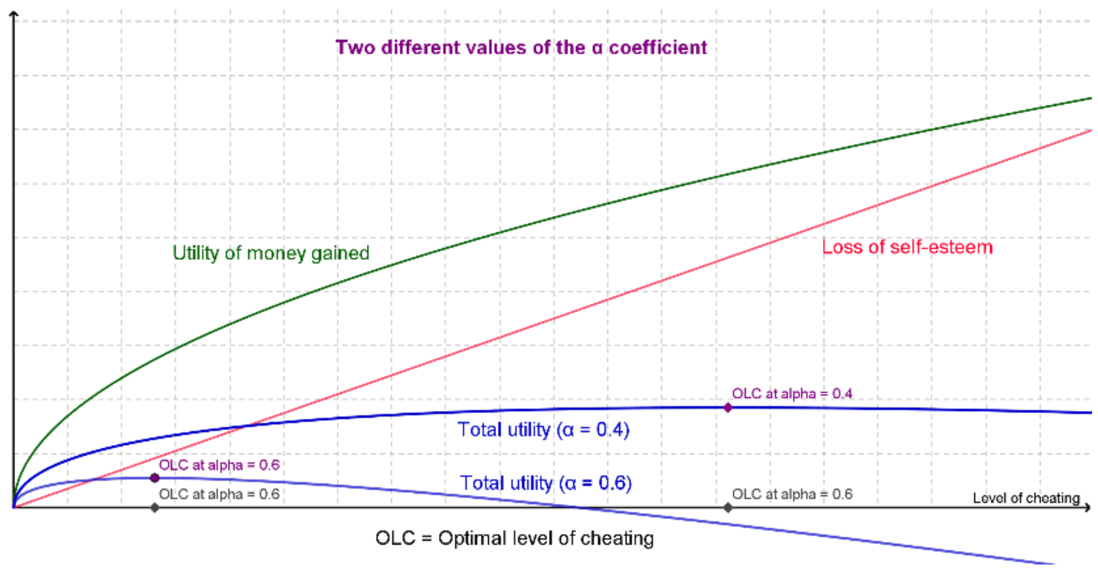


Figure 5. Total utility as the weighted sum of the utility of money gained and the disutility of loss of self-esteem for two levels of α.

With such functions, a DM would decide to cheat for incentives of magnitude *x* which give *U (x)* > 0, refrain from cheating for those with *U (x)* < 0, and their willingness to cheat would be highest at the maximum of the *U* function, denoted as the "optimal level of cheating". Both the range of preference for cheating $U^{-1}(\mathrm{R}_{+})$ and the optimal level of cheating will depend on the coefficient α; in Figure 5 this is depicted for two different values of α.

The parameter α is a number ranging from 0 in the case of no concern for the moral values involved in a decision (a purely "homo oeconomicus" DM) to 1 for a person that would never cheat. It corresponds to the level of ambiguity of the decision situation in our study: the higher the ambiguity, the weaker a DM's inclination to consider the moral aspect of their decision. It can be easily seen from Figure 5 that for higher values of α both the optimal level of cheating and the range of *x* within which a DM decides to cheat will be smaller than for lower values. Also, assuming the utility function $u_s$ suggested by prospect theory and predicting greater utility for avoiding a loss *x* than for receiving a gain *x*, the model immediately shows a stronger willingness to cheat for the same stakes under loss framing (with a steeper $u_s$) than under gain framing.

Comprehensive testing of such a model, including its additive form and estimating the parameters of utility functions, is clearly a complex task far beyond the scope of this work. Even so, under plausible and realistic assumptions made about the utility functions, the model can explain a rich variety of effects that have been observed in both real life and in numerous experimental studies, including those confirmed by the experiment reported in this paper. For example, the model can predict the behavior of a person with an S-shaped $u_s$ function (see Figure 6) who badly needs a certain amount of money (e.g., for medical treatment as in the Heinz dilemma; (Kohlberg, 1981) to save their life, or to save the reputation of a beloved person as in Dostoevsky's novel The Gambler (Dostoyevsky, 1996, 2009).

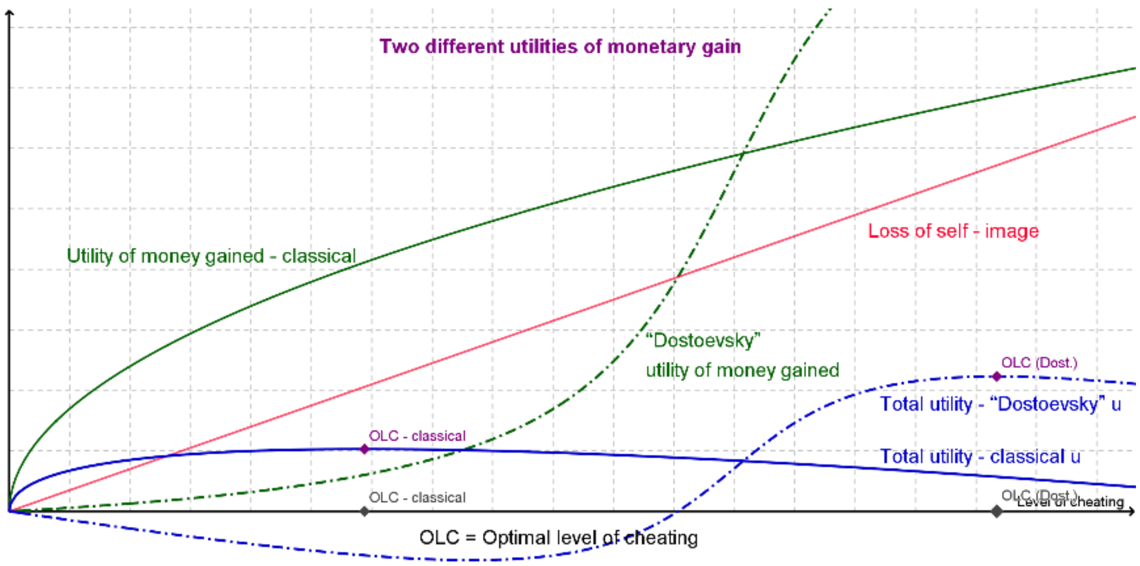


Figure 6. Total utility as the sum of the utility of money gained and the disutility of loss of self-esteem: two different $u_s$ functions

One can also compare people with linear vs. convex loss of self-esteem disutility functions as presented in Figure 7. A person with a convex disutility function should cheat less than a person with a linear $u_m(x)$ loss of self-esteem disutility function. In fact, a person with a convex disutility function can follow the prediction of Mazar et al. (2008) relatively easily and cheat a little bit with no significant drop in self-esteem, but the moral costs rise quickly with increasing magnitude of cheating.

Both the above points are examples of how our model makes different predictions based on the assumed shape of both functions.

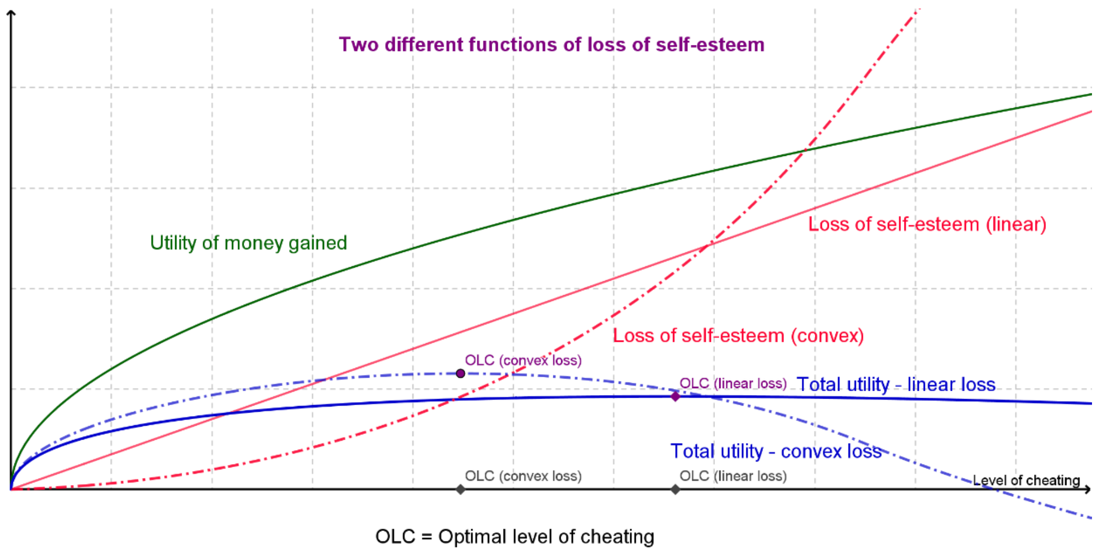


Figure 7. Total utility as the sum of the utility of money gained and two disutility functions relating to loss of self-esteem.

Generally, humans' decisions in situations involving moral aspects could be predicted and compared with the model if we had some knowledge of their utility functions, in particular knowledge about disutility $u_m$. However, we believe that decision making processes in such situations involve multiple criteria and tradeoffs similar to those encountered in purely economic decisions, and that these processes obey similar principles.

**ACKNOWLEDGEMENT**

This work was supported by the National Science Centre, Poland (NCN) under Grant 2015/17/D/HS6/02684.

## REFERENCES


Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica, 87*(4), 1115-1153. doi:10.3982/ecta14673

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(i01), 1-48. doi:doi:10.18637/jss.v067.i01

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(2), 169-217. doi:10.1086/259394

Bereby-Meyer, Y., Hayakawa, S., Shalvi, S., Corey, J., Costa, A., & Keysar, B. (2018). Honesty speaks a second language. *Topics in Cognitive Science*, 1-12. doi:DOI: 10.1111/tops.12360

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), 378-400.

Chandler, J. J., & Paolacci, G. (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological and Personality Science, 8*(5), 500-508. doi:10.1177/1948550617698203

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1-8. doi:https://doi.org/10.1016/j.jebo.2011.08.009

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412-431. doi:https://doi.org/10.1016/0001-6918(69)90065-1

Dostoyevsky, F. (1996). *The Gambler*: Dover Publications.

Dostoyevsky, F. (2009). *Dostoevsky: Letters and Reminiscences*: Kessinger Publishing Company.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. doi:10.3758/brm.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. doi:10.3758/BF03193146

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

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

Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological Science, 21*(5), 712-720. doi:10.1177/0956797610366545

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197-216. doi:10.1037/0022-3514.85.2.197

Grolleau, G., Kocher, M. G., & Sutan, A. (2016). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science, 62*(12), 3428-3438. doi:10.1287/mnsc.2015.2313

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition*, *163*, 15-25.

Hochman, G., Glöckner, A., Fiedler, S., & Ayal, S. (2016). "I can see it in your eyes": Biased Processing and Increased Arousal in Dishonest Responses. *Journal of Behavioral Decision Making*, *29*(2-3), 322-335. doi:10.1002/bdm.1932

Inquisit. (2015). Inquisit: Computer software (Version 4). Retrieved from http://www.millisecond.com.

Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2017). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys, 32*(2). doi:10.1111/joes.12204

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi:https://doi.org/10.1016/j.jml.2007.11.007

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*: Elsevier.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi:10.1037/a0028347

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior, 102*(Supplement C), 433-444. doi:10.1016/j.geb.2017.01.015

Kohlberg, L. (1981). *Essays on moral development: The psychology of moral development* (Vol. 2): San Francisco: harper & row.

Leib, M., Pittarello, A., Gordon-Hecker, T., Shalvi, S., & Roskes, M. (2019). Loss framing increases self-serving mistakes (but does not alter attention). *Journal of Experimental Social Psychology*, *85*, 103880. doi:https://doi.org/10.1016/j.jesp.2019.103880

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433-442. doi:10.3758/s13428-016-0727-z

Markiewicz, Ł., & Czupryna, M. (2019). Cheating: One common morality for gains and losses but two components of morality itself. *Journal of Behavioral Decision Making, 0*(0). doi:10.1002/bdm.2151

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

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

Mazar, N., & Zhong, C.-B. (2010). Do green products make us better people? *Psychological Science*, *21*(4), 494-498. doi:10.1177/0956797610363538

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

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023-1031. doi:10.3758/s13428-013-0434-y

Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical blind spots. *Psychological Science, 26*(6), 794-804. doi:10.1177/0956797615571018

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Rahwan, Z., Hauser, O. P., Kochanowska, E., & Fasolo, B. (2018). High stakes: A little more cheating, a lot less charity. *Journal of Economic Behavior & Organization, 152*, 276-295. doi:https://doi.org/10.1016/j.jebo.2018.04.021

Rosenbaum, S. M., Billinger, S., & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology, 45*(Supplement C), 181-196. doi:10.1016/j.joep.2014.10.002

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172-177. doi:10.1016/j.jesp.2016.09.009

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

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin, 143*(4), 428-453. doi:10.1037/bul0000087

Zhong, C., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science, 21*(3), 311-314. doi:10.1177/0956797609360754