Term paper

# Honesty nudges: classroom experiment on priming vs. moral licensing

Matej Lorko

**Abstract**: Emotional priming and moral licensing conjectures diverge in their predictions regarding the efficacy of honesty nudges. Priming theory assumes that if individual is nudged to recall a situation in which she behaved honestly, she continues to behave honestly in the future. On the other hand, moral licensing theory assumes that after recalling one's own honest behavior, the individual might reconcile his actions by behaving more dishonestly in the future. I run a classroom experiment to test the effect of a nudge to recall either honest or dishonest past behavior on a decision in a simple task in which it is favorable for most individuals to cheat. I find null effects of honesty nudges, providing no support for priming conjecture nor moral licensing conjecture. However, the result might be driven by the fact that participants in the experiment tended to cheat very moderately, if at all.

## 1. Introduction

Nudges are as "aspects of the choices architecture that alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008, p.6). Given that nudging interventions are cheap and easy to implement, nudges have gained a lot of interest by not only academics, but also by government officials and business professionals. DellaVigna and Linos (2022) estimate that there are over 200 government "nudge units" all over the world, implementing behavioral insights and nudges into public policies. Notable companies employing and leveraging nudges at scale include Uber, Google, Amazon, and Walmart.

One of the most prominent types of nudges are honesty nudges, i.e., psychological interventions aiming to increase the individual honesty or compliance with social and other norms. For example, in the context of tax compliance, a rich variety of nudges have been applied, ranging from appeals to peer examples (Hallsworth et al., 2017), highlighting public goods provision (Bott et al., 2020), to correcting procrastination and limited attention (Mascagni & Nell, 2022; Hernandez et al., 2017; Loewenstein & Wojtowicz, 2023).

However, there has been a lot of controversy regarding the replicability of experiments on honesty nudges. For example, while Mazar, Amir & Ariely (2008) report that people who were given the opportunity to cheat did not cheat if they were first asked to recall the 10 commandments, a follow up replication study reports null effects of 10 commandments recall over 25 individual experiments (Verschuere et. al., 2018). Similarly, a large-scale replication study of Kristal et. al. (2020) refutes any positive effects of signing at the beginning of the form instead of at the end, contrary to results of original study of Shu et. al. (2012). Even more strikingly, the Shu et. al. (2012) study was recently retracted due to allegations of data fabrication.

In this study, I ran a classroom experiment in which students individually and privately roll a dice and then self-report their outcome. Given that higher outcome earns them more points towards their final grade, they have incentives to cheat. Students are randomly assigned into a Baseline treatment, a Honesty treatment in which, before reporting their dice roll, they are asked to recall a situation from their life in which they had an opportunity to cheat or lie but decided to behave honestly, and a Dishonesty treatment in which they are asked to recall a situation in which they had an opportunity to cheat or lie and decided to take it.

Interestingly, research in psychology offers two directionally opposite predictions of the honesty nudge effects. While emotional priming conjecture (see e.g., Kahneman, 2011) assumes that nudging to recall honest behavior results in subsequent honest behavior, moral licensing conjecture (see e.g., Blanken et. al., 2015) assumes that if an individual is nudged to recall a past honest behavior, she might issue herself a "moral license" for subsequent dishonesty. Therefore, I derive the following contradictory hypotheses:

- **Hypothesis 1a** (based on emotional priming): Honesty recall nudge results in a more honest behavior.
- **Hypothesis 1b** (based on moral licensing): Honesty recall nudge results in a more dishonest behavior.

Given its definition, emotional priming should affect behavior irrespectively on the nature of the nudge. That said, I conjecture that dishonesty recall nudge results in more dishonest behavior. On the other hand, moral licensing hypothesis is silent when it comes to the effect of dishonesty nudge. Therefore, I derive only one hypothesis related to the effects of dishonesty nudge.

- **Hypothesis 2** (based on emotional priming): Dishonesty recall nudge results in more dishonest behavior.

## 2. Experimental design

The experiment consists of two steps. In the first step, each subject privately draws a random integer number between one and six, effectively simulating a standard six-sided dice roll. In the second step, subject privately fills an online form eliciting her name, age, gender, study group, red diploma aspirations and outcome of the dice roll. A subject is informed that a roll of 1 or 2 will earn her 1 point towards the final grade, a roll of 3 or 4 will earn her 2 points and a roll of 5 or 6 will earn her 3 points.

Depending on the assigned treatment, subject might be requested to recall her past behavior before self-reporting her dice roll. I implement three treatments. The Baseline treatment omits the recall task. In the Honesty treatment, subject is nudged to recall a past situation in which she behaved honestly, while in the Dishonesty treatment, subject is nudged to recall a situation in which she behaved dishonestly. The three treatments along with exact wording of the nudge are summarized in Table 1.

**Table 1: Treatment overview**

| Treatment | Nudge |
| --- | --- |
| Baseline | No nudge |
| Honesty | Now recall a situation from your life, when you had an opportunity to lie or cheat but you decided not to take it. That is, a situation in which you decided to behave honestly. Please take at least a minute of your time to recall that situation in as much detail as possible. |
| Dishonesty | Now recall a situation from your life, when you had an opportunity to lie or cheat and you decided to take it. That is, a situation in which you decided to behave dishonestly. Please take at least a minute of your time to recall that situation in as much detail as possible. |

### 3. Results

A total of 102 subjects, with a mean age of 22.3 years (standard deviation 2.3 years) and 54% female, participated in the experiment. The number of subjects in each treatment ranged from 30 to 39. All subjects were students at the University of Economics in Bratislava, attending a lecture of Introduction to experimental methods in economics course.

**Table 2: Mean dice rolls (SD) by treatments and study groups**

| Treatment | N | Mean roll | Group | N | Mean roll |
|---|---|---|---|---|---|
| Baseline | 33 | 3.7 (1.6) | Slovak – in person | 80 | 3.4 (1.5) |
| Honesty | 39 | 3.8 (1.5) | Slovak - online | 11 | 2.5 (1.9) |
| Dishonesty | 30 | 2.6 (1.7) | English | 11 | 4.5 (1.7) |

Table 2 presents the mean dice rolls and respective standard deviations by treatments (left panel) and by study groups (right panel). As is apparent from the table (see also Figure 1), subjects generally reported their dice rolls truthfully, as the average roll off all subjects yields 3.4, which is not statistically significantly different from a theoretical expected average of 3.5 (t-test, p-value = 0.67). In addition, I find no significant differences in dice rolls between the Baseline treatment and Honesty treatment (t-test, p-value = 0.80), yielding no support for my Hypotheses 1a nor 1b.

***Result 1:*** *Honesty nudge does not affect moral behavior.*

**Figure 1: Mean dice rolls by treatment**
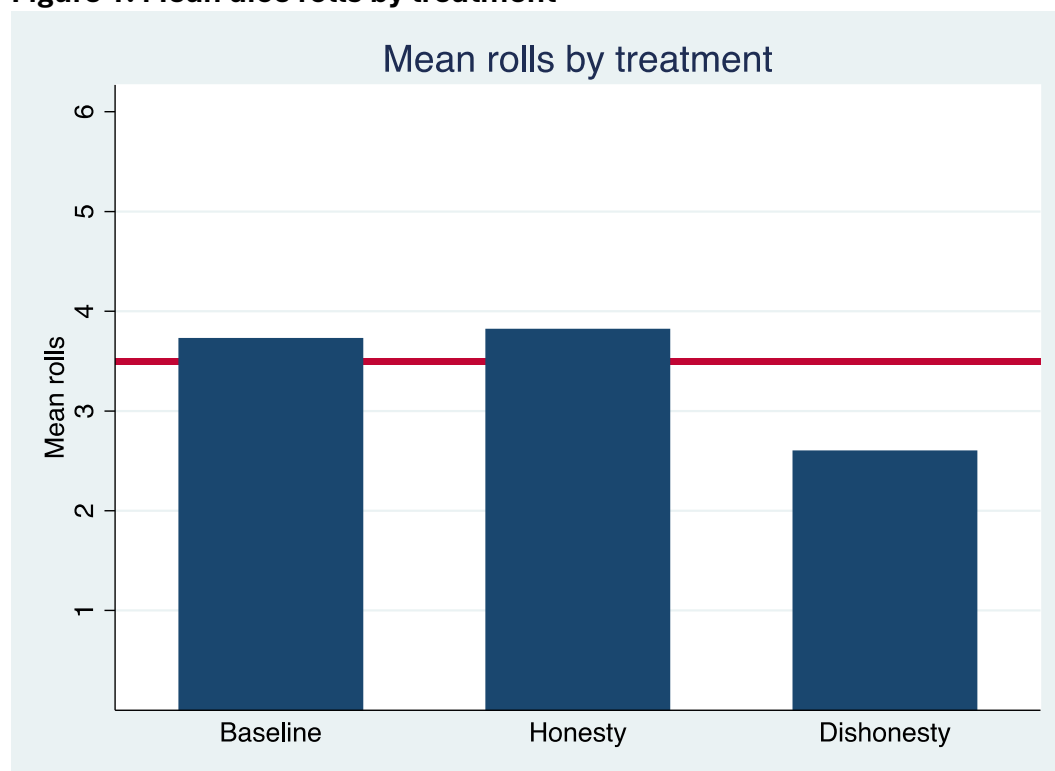


Note: For reference, the red line indicates a theoretical expected mean of 3.5.

Interestingly, the mean dice roll is the lowest in the Dishonesty treatment. Pairwise comparison of mean dice rolls between the Dishonesty treatment and the Baseline and Honesty treatments shows that the differences are statistically significant (t-test, p-values = 0.01 and 0.002 respectively). However, given that the mean dice roll in Dishonesty treatment is lower compared to the expected mean, the differences cannot be attributed to the effect of nudges. Rather, they are most likely driven simply by unluckier rolls in the Dishonesty treatment. Therefore, I find no support for my Hypothesis 2.

**Result 2:** *Dishonesty nudge does not affect moral behavior.*

I check the robustness of my findings using OLS regressions, presented in Table 3. In Model 1 I only use dummy variables for treatments, in Model 2 I add controls for age, gender and red diploma aspirations and in Model 3 I add controls for the study groups. I find no significant associations between basic demographics (age, gender), aspirations and dice rolls. However, in Model 3 I find that online students report weakly significantly lower dice roll outcomes (most probably due to unluckier draws), while students in English group report weakly significantly higher outcomes, which can be explained either by better luck or by higher tendency to cheat.

### Table 3: OLS regressions

|  | (1) Dice roll | (2) Dice roll | (3) Dice roll |
|---|---|---|---|
| **Honesty treatment** | 0.09 (0.36) | 0.13 (0.37) | 0.09 (0.36) |
| **Dishonesty treatment** | -1.13*** (0.41) | -1.10** (0.42) | -0.97** (0.41) |
| **Age** |  | -0.04 (0.10) | 0.06 (0.11) |
| **Female** |  | 0.09 (0.33) | 0.09 (0.32) |
| **Goes for red diploma** |  | -0.15 (0.45) | -0.06 (0.46) |
| **Slovak online group** |  |  | -1.02* (0.55) |
| **English group** |  |  | 1.04* (0.57) |
| **Constant** | 3.73*** (0.28) | 4.60* (2.32) | 2.39 (2.41) |
| **N** | 102 | 102 | 102 |
| **R²** | 0.11 | 0.11 | 0.18 |

Notes: Standard errors are reported in parentheses.
*, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

## 4. Discussion

While a considerable number of earlier studies find positive effects of honesty nudges due to emotional priming, more recent replication studies seem to refute those findings. In line with growing number of unsuccessful replications, I find no apparent effects of honesty nudges. In addition, I also find no support for moral licensing hypothesis.

However, it is important to note that in the current experiment, the room for finding effects of honesty nudges was very limited due to negligible magnitudes of cheating in general. In fact, based on the results, it is very likely that vast majority of participants decided to report their outcomes truthfully irrespective of their treatment assignment. While I find weakly significantly higher mean dice roll for students in English group, due to low number of observations (only 11 students), I cannot reject that this result was driven by luck instead of dishonesty. Future research could explore the effects honesty nudging in less conservative setting, perhaps by providing stronger incentives to cheat.

## References

Blanken, I., Van De Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540-558.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science, 66*(7), 2801–2819.

DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, *90*(1), 81–116.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31.

Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). *Applying Behavioral Insights to Improve Tax Collection*.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan

Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. Proceedings of the National Academy of Sciences, 117(13), 7103-7107.

Loewenstein, G., & Wojtowicz, Z. (2023). The Economics of Attention. *Available at SSRN 4368304*.

Mascagni, G., & Nell, C. (2022). Tax compliance in Rwanda: Evidence from a message field experiment. *Economic Development and Cultural Change*, *70*(2), 587–623.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of selfconcept maintenance. Journal of marketing research, 45(6), 633-644.

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., ... & Yıldız, E. (2018). Registered replication report on Mazar, Amir, and Ariely (2008). Advances in Methods and Practices in Psychological Science, 1(3), 299-317.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proceedings of the National Academy of Sciences, 109(38), 15197-15200

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.