Delayed Honesty*

Pre-Analyis Plan

Stefania Bortolotti   Felix Kölle   Lukas Wenner

November 2018

1   Introduction

Many economic interaction involve the release of private information to other agents. Often, there is a trade-off between truthful reporting, which may be morally desirable and/or socially optimal and lying, which may be individually optimal, e.g., payoff-maximizing. Rejecting the idea that economic agents are purely interested in their own monetary gains and without any "moral compass", a recent literature in economics has demonstrated a number of factors which influence an individual's degree of honesty. These factors include, among others, the possible gains from lying, the degree of social observability, and the ex-ante likelihood of different states of the world.

A (to our knowledge) completely unexplored area revolves around how the relative timing of dishonest behavior and the realization of benefits, affect people's degree of dishonesty. Overcoming the restrictive nature of previous work where benefits from dishonest reporting always accrue (almost) immediately allows us to study novel factors which may affect the degree of honesty in society. In particular, we are interested in the following three questions: (i) How is lying affected when the benefits from misreporting accrue delayed rather than immediately? (ii) How does increased mental engagement with the lie affect misreporting when benefits are delayed? (iii) How does a delay between the time when private information is obtained and the time when people report it affect dishonesty?

To answer these questions, we plan to conduct the following experiment.

2   Experimental Design

Our design consists of six treatments. In all treatments, subjects participate in a study that consists of two separate experimental sessions, taking place exactly one week apart from each other. In the following, we will refer to these sessions as week 1 and week 2. The experiment will be conducted online, and subjects are recruited via the online labor market platform MTurk.

In each session, participants first rate a series of 20 paintings. For that, they are always shown two different paintings side-by-side and have to indicate which of the two they like better. We emphasize that we are interested in their subjective opinion and that there is no right or wrong answer. The purpose of this task is to not reveal the main aim of this study and to make sure that in all treatments there is a meaningful task in both weeks. After that, they move on to the main part of the experiment. There, they are being told that they can earn some additional money (a bonus), which will depend on a coin flip task. The exact nature of the coin flip task thereby varied across treatments, as described below.

2.1   Treatments

In all of our treatments, subjects are asked to flip a coin twice in private, and to report the outcome of the coin flip to us. This report then determines the bonus payment paid to the subject; if the subject reports 2 Heads, she receives a bonus payment of $2.00, whereas any other report (1 Heads and 1 Tails or 2 Tails) yields a bonus of $0.25. While the rule is that subjects have to report the outcome of the coin flip truthfully, since we cannot observe the actual outcome, subjects may choose to misreport in order to achieve a higher bonus. Across our different treatments, we vary when this bonus will be paid (either immediately or delayed), the degree to which subjects will be reminded of their potentially untruthful behavior, as well as when the outcome of the two coin flips has to be reported.

- *Immediate*: Subjects report the outcome of the coin flips in week 1, and receive their corresponding bonus on the same day.

- *PayDelay*: Subjects report the outcome of the coin flips in week 1, and receive their corresponding bonus with a delay of one week, just before the start of the second session.

- *PayDelay+ShowUp*: Subjects report the outcome of the coin flips in week 1, and receive their corresponding bonus in week 2, after they completed the second session. They

receive this payment only if they take part in the second session.

- *PayDelay+Recall*: Subjects report the outcome of the coin flips in week 1. They receive their corresponding bonus in week 2, but only if they can correctly recall their report from week 1. If their reports from both weeks do not match, they receive no bonus.

- *ReportDelay*: Subjects are instructed to flip the coin in week 1 but only report the outcome of the coin flip in week 2. They receive their corresponding bonus after they completed the second session.

- *Immediate-Week2*: Same as *Immediate* with the exception that the coin flip task now takes place in week 2 instead of week 1.

2.2 Procedures

A total of 1240 workers will be recruited via MTurk and randomly assigned to one of the six treatments. Participation will be restricted to workers residing in the US and with a high approval rate (90% or more). The study itself will be implemented with Qualtrics. In week 1, all participants, expect participants in treatment *ImmediateWeek2* will have to answer a set of control questions about the coin flip task. Only participants who successfully answer all control questions will be allowed to complete the study.[1]

In week 2, only workers who successfully completed the study in week 1 will be allowed to participate. This will be clearly announced in week 1, where participants will be made aware of the rules for both weeks. To minimize attrition from week 1 to week 2, we will send an email to everybody who took part in the first session and remind them of the second part of the experiment. We send this email seven days after the first experiment (in the morning). After receiving the email, subjects have 24 hours to complete the second part of the experiment. In week 2, subjects in treatments *ReportDelay* and *ImmediateWeek2* will have to answer the control questions and only be allowed to continue with the study if they answer these correctly.

For the successful completion of each of the two HITs, participants receive a flat payment of $0.50, paid to them directly after completing the HIT. To further limit attrition across weeks, participants receive a completion payment of $1.50 if they successfully complete both HITs. The completion payment is paid to them three days after the second session. Table 1 provides an overview of our experimental design.

---

[1]After two incorrect answers, participants will be automatically excluded and will be prevented from re-taking the study.

Table 1: Timeline of the experiment

	HIT 1 (2018/11/27)				HIT 2 (2018/12/04)				2018/12/07
	10am (ET)		from 10am onwards	4am (ET)	6am (ET)	6pm (ET)		from 6am onwards	4am (ET)
Immediate	invite	flip + report	base ($0.50) + bonus		invite	reminder	—	base ($0.50)	completion ($1.50)
PayDelay	invite	flip + report	base ($0.50)	bonus	invite	reminder	—	base ($0.50)	completion ($1.50)
PayDelay+ShowUp	invite	flip + report	base ($0.50)		invite	reminder	—	base ($0.50) + bonus	completion ($1.50)
PayDelay+Recall	invite	flip + report	base ($0.50)		invite	reminder	recall	base ($0.50) + bonus	completion ($1.50)
ReportDelay	invite	flip	base ($0.50)		invite	reminder	report	base ($0.50) + bonus	completion ($1.50)
Immediate-Week2	invite	—	base ($0.50)		invite	reminder	flip + report	base ($0.50) + bonus	completion ($1.50)

Note: "—" indicates that no action with respect to the bonus needs to be taken. However, if a subject does not participate in this stage, he does not receive any of the future payments.

2.3 Sample Size & Pilot Study

We plan to obtain 200 data points per treatment. Our variable of interest is the proportion of subjects who report *2 Heads*. Power calculations based on a z-test for the difference of independent proportions reveal that we have a power of 80% to reject the equality of proportions if the normalized effect size h is above 0.28. Depending on the "baseline", i.e., the absolute proportion of reports of *2 Heads*, this corresponds to a difference of 0.13 to 0.15 for lying rates which are within the range of previous studies.

In early November 2018, we conducted a pilot study to investigate whether our incentives are strong enough to ensure that participants would return for HIT 2. We therefore implemented an experiment which was very similar to the current experimental design with respect to the timing, but where participants did not have the possibility to behave dishonestly. We found an attrition rate of 10% which we found low enough to proceed as planned. By design, no data on lying behavior was obtained. The attrition rate of about 10% is why for treatments *ReportDelay* and *ImmediateWeek2* we recruit 220 participants in total, because there we will use the week 2 data for analysis (see below).

3 Predictions

We seek to test the following three predictions.

Prediction 1 (Delayed Payments). *The proportion of reports of 2 Heads is higher in Immediate than in PayDelay.*

Intuition: The two treatments are identical, apart from the date on which the bonus is paid. Under the assumption that some agents face a fixed cost of lying, and that this cost realizes at the time of the untruthful report, we predict that when the bonus is paid later, this increases honesty. The reason is that relative to *Immediate* in *PayDelay* the gains from lying are smaller because agents discount future payments, while the costs of lying occur in the present are thus the same in both cases.

Prediction 2 (Engagement Costs of Dishonesty). *The proportion of reports of 2 Heads is higher in PayDelay+ShowUp than in PayDelay+Recall.*

Intuition: The two treatments are identical, apart from the degree to which an agent "engages" with the report during the seven days between the two treatments. While in both cases participants need to complete HIT 2 in order to receive the bonus payment from the

coin flip task, in *PayDelay+ShowUp* they do not need to engage with the task between the two HITs. In *PayDelay+Recall*, in contrast, in HIT 2 participants need to correctly recall what they reported in HIT 1. In case they reported untruthfully, they thus have to remember this during the whole week. Anticipating that being reminded of having behaved dishonestly comes at an increased cost (e.g., reduced self image), agents may report honestly instead.

Prediction 3 (**Motivated Recall**). *The proportion of reports of 2 Heads is higher in ReportDelay than in ImmediateWeek2.*

Intuition: The two treatments are identical, apart from the time when the coin is flipped. While in *ImmediateWeek2* the coin flip and the report both happen in week 2, in *ReportDelay* the coin is flipped in week 1 but the outcome of the coin flip has to be reported only in week 2. In the latter case, an agent may find it "convenient" to forget the outcome of the coin flip because if she is unable to recall, or is unsure about what the exact outcome was, reporting *2 Heads* might come at a lower cost as agents can maintain their self-image of being honest.

4   Empirical Strategy

To test each of the three hypotheses, a standard χ^2-test is employed, comparing the number of times a report of 2 Heads is made. From this, the share of dishonest people can be (approximately) calculated. Since the expected number of 2 Heads is 25%, the difference between the actual proportion of reports of 2 Heads in the sample and 25% is the share of liars in a given treatment.

It is important to decide ex-ante how to deal with the possibility of attrition in our study. Even though the completion payment of $1.50 and the generally high incentives for an MTurk study are chosen to minimize attrition, this cannot be completely ruled out. We decided to use different exclusion criteria, depending on the hypothesis to be tested.

For the test of Prediction 1, we use the full sample of people who completed the study in **week 1** since there is no relevant decision to be made in week 2, and the bonus is paid regardless of whether subjects complete the study or not.

For the test of Prediction 2, we use the sample of people who completed the study in **week 1**. We decided to do so, because reasons for attrition may be different in the two treatments. In particular, subjects in *PayDelay+Recall* may forget what they reported and therefore not return. Moreover, if forgetting is not anticipated but correlated with the reported outcome in week 1, then using the week 2 data would bias our results.

For the test of Prediction 3, we only include people who actually return and participate

in **week 2** as this is the week where these people make their report. For participants who do not return in week 2, we have no data to analyze.