Truth-Telling and the Regulator

Pre-analysis plan for a modified conceptual replication with Brexit Leave voters

Moritz A. Drupp[a], Menusch Khadjavi[b,c], Martin F. Quaas[c,d,e]

[a] *Department of Economics, University of Hamburg, Germany*

[b] *Department of Economics, University of Kiel, Germany*

[c] *Kiel Institute of the World Economy, Germany*

[d] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany*

[e] *Department of Economic, University of Leipzig, Germany*

March 06, 2019

This is a pre-analysis plan for a modified conceptual replication study of an artefactual mail field experiment with all commercial German fishermen on how truth-telling, as measured by a standard coin-tossing task, depends on the salience of their ill-regarded regulator, the EU, that we ran in winter 2015/2016.[1] We used the 4-coin toss task of Abeler et al. (2014) and varied the salience of the regulator by varying the letterhead. We sent out three versions of the instructions in a between-subjects design: (i) a baseline setting ('Baseline') in which only the logos of university institution 1 and institution 2 present on the letterhead, (ii) a version where the EU flag is made salient in the letterhead of the instruction sheet ('EU_Flag'), and (iii) an additional treatment where the framing states that this research has been funded by the European Commission ('EU_Flag_Funding').

---

[1] A previous version of this study is available as a discussion paper at: https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/Menusch_Khadjavi/truth-telling-and-the-regulator-evidence-from-a-field-experiment-with-commercial-fishermen/kwp-2063.pdf.
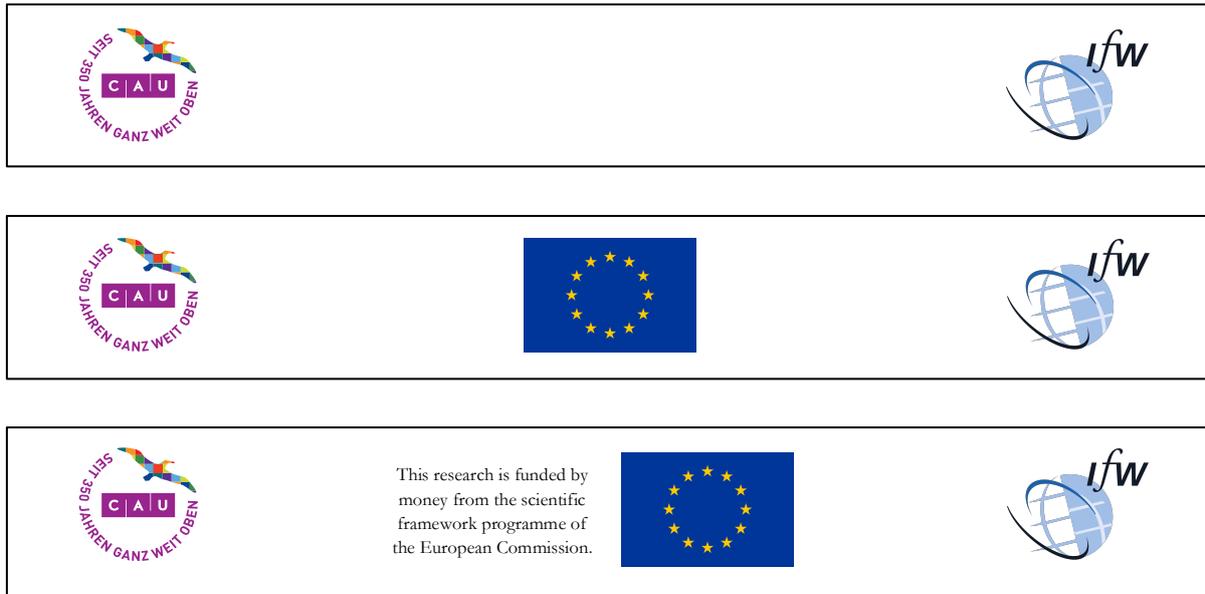
Figure 1: Letterheads of the three treatments (from top to bottom: Baseline, EU_Flag and EU_Flag_Funding).

We tested the following hypotheses:

HYPOTHESIS 1: Fishermen report greater tail-tosses than the truthful distribution, but do not fully misreport in the Baseline treatment.

HYPOTHESIS 2: Fishermen report less truthfully in the EU_Flag treatment compared to the Baseline treatment.

HYPOTHESIS 3: Fishermen report even less truthfully in EU_Flag_Funding compared to the EU_Flag treatment.

Only 120 of the 896 commercial fishermen participated in our experiment, which is why we ended up with a smaller than expected sample size in each experimental cell. The purpose of the current study is to conceptually replicate the main result of the first study, which provides evidence supporting Hypothesis 2, and to provide further insights into Hypothesis 3. Contrary to that hypothesis, we found that some fishermen were more honest in the treatment where the source of EU research funding is revealed. It may be that the information on research funding leads to more honest reporting, or that it induces a (stronger) experimenter demand effect. To disentangle the effect of the EU_Flag and the funding information per se, we modify the third treatment, that is letterhead, in the replication study. Instead of featuring both the EU flag and

the EU funding information, it now features "This study is funded by public research funds." We denote this treatment 'Funding'. The new screen headers are depicted in Figure 2.
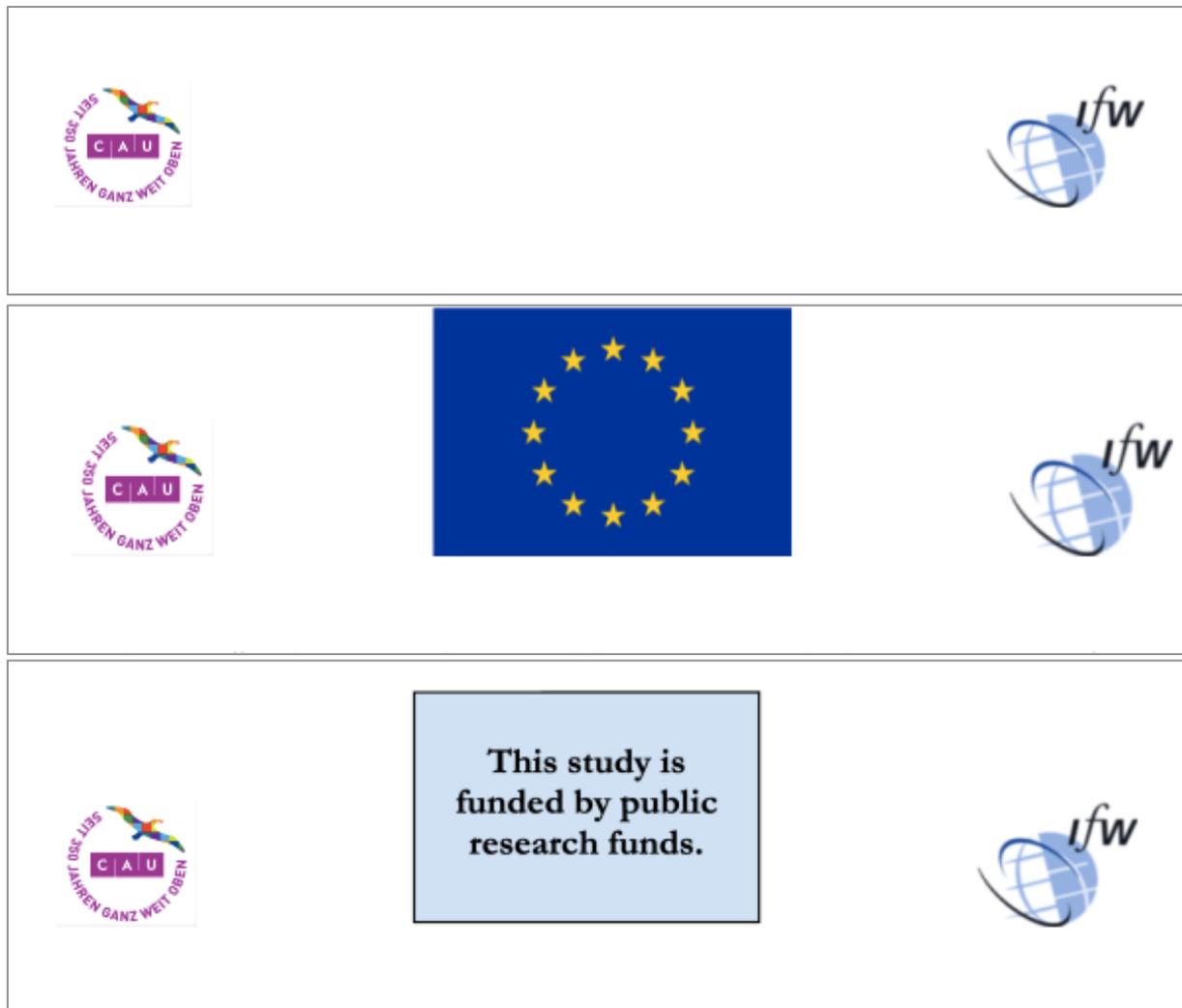
Figure 2: Screen headers of the three treatments (from top to bottom: Baseline, EU_Flag and Funding).

For our purposes, we need a sample of respondents that is as similar to the fishermen with respect to the (negative) stance towards the EU as possible. We have identified Brexit voters as a candidate population. We will run our conceptual replication via the platform www.prolific.ac. They have numerous pre-screening options based on how participants have answered to previous questionnaires and one of them is how people have voted in the Brexit referendum. We will restrict participation in our study to those who have previously noted that they voted for LEAVE, and who have a high approval rate of 95 of 100 in previous studies on Prolific (this rate is an indicator for the quality of participants' answers, e.g. to avoid very fast-clicking, sloppy

participants etc.). This narrows down the potential participant population in Prolific to 4064 subjects.

We will nevertheless both inquire again how participants have voted in the referendum and will also ask how they would vote today.[2] We will pay participants 2 GBP for answering a series of questions (see below) and additionally a variable payment for completing the coin toss task (0 to 4 GBP, i.e. 1 GBP per "successful" coin toss). Given our budget constraint, we will invite a maximum number of 1200 participants, which will be randomly allocated to one of the three treatments by the software of the survey tool www.soscisurvey.de. Due to a deadline to submit the revised version of our manuscript, we will stop the survey on Wednesday, March 13[th,] even if this maximum quota has not been reached yet.[3]

Table 1 reports a list of all variables that we will collect on the relevant screens (note that Screen 1 is a welcome screen without data collection). Overall, we will collect 16 variables.

Table 1. List of all Variables to be Collected.

Variable # | type	Screen 2: Pre-questions
1 | numeric	age
2 | categorical	gender
3 | numeric	relative_income
4 | numeric	moves
5 | numeric	trust_local_gov
6 | numeric	trust_royal_society
7 | numeric	trust_UK_gov
8 | numeric	trust_EU
	Screen 3: Abeler et al. (2014) truth-telling task (flipping a coin four times)
9 | numeric	Variable name: [coinreport]
	Screens 4 to end: Follow-up questions
10 | numeric	Variable name: [exp_demand] How strongly do you feel that the researchers of this study wanted you to report in a particular way in the coin task?
11 | numeric	Variable name: [seconds] How long did it take for you to fetch the coin to toss?
12 | numeric	Variable name: [revenge]

---

[2] We have conducted a small pre-test in prolific with six REMAIN voters and found that indeed six had voted REMAIN and still five of the six would do so again.
[3] If only a small number of participants have taken part and if we were granted a deadline extension, we would aim for a second wave to fill the 1200 data slots.

	If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.
13 | numeric	Variable name: [punish] How willing are you to punish someone who treats you unfairly, even if there may be costs for you?
14 | numeric	Variable name: [revenge_others] How willing are you to punish someone who treats others unfairly, even if there may be costs for you?
15 | categorical	Variable name: [Brexit] How did you vote in the Brexit referendum?
16 | categorical	Variable name: [Brexit2] How would you vote if the referendum on the Brexit would have been taken now?

In the following, we will present statistical tests that we will run. As described above, we have three treatments in this modified conceptual replication: 'Baseline', 'EU_Flag' and 'Funding'. Our main outcome variable is truth-telling as measured by variable 9 [coinreport].

Hypothesis R1a: The average number of reported tail tosses is greater in EU_Flag compared to Baseline. We will test this hypothesis by means of a (one-sided) t-test.

Hypothesis R1b: The frequency of subjects reporting 4 tail tosses in EU_Flag is greater than in Baseline. We will test this hypothesis by means of a chi-squared test.

Hypothesis R1c: The frequency of subjects reporting 1 and 2 tail tosses in EU_Flag is lower than in Baseline. We will test this hypothesis by means of a chi-squared test.

Hypothesis R2a: The average number of reported tail tosses is greater in EU_Flag compared to Funding. We will test this hypothesis by means of a (one-sided) t-test.

Hypothesis R2b: The frequency of subjects reporting 4 tail tosses in EU_Flag is greater than in Funding. We will test this hypothesis by means of a chi-squared test.

Hypothesis R2c: The frequency of subjects reporting 1 and 2 tail tosses in EU_Flag is lower than in Funding. We will test this hypothesis by means of a chi-squared test.

Hypothesis R3a: The average number of reported tail tosses is greater in Baseline compared to Funding. We will test this hypothesis by means of a (one-sided) t-test.

Hypothesis R3b: The frequency of subjects reporting 4 tail tosses in Baseline is greater than in Funding. We will test this hypothesis by means of a chi-squared test.

Hypothesis R3c: The frequency of subjects reporting 1 and 2 tail tosses in Baseline is lower than in Funding. We will test this hypothesis by means of a chi-squared test.

We will complement our treatment effect analyses with a regression analysis that includes further variables that aim to reveal mechanisms of our expected effects and control variables.

We will collect data on the level of trust in the EU on a Likert scale (variable 8 in Table 1 above):

Hypothesis R4a: There is a negative correlation between the number of reported tail tosses in EU_Flag and the magnitude of trust in the EU. That is, the more a subject trusts the EU, the lower will be her/his number of reported tail tosses in EU_Flag.

Hypothesis R4b: There is a relatively less strong negative (or no significant) correlation between the number of reported tail tosses in Baseline and Funding and the magnitude of trust in the EU.

Furthermore, we will test whether there is an indication of experimenter demand effects that might differ across treatments. We will measure this by variable 10 [exp_demand]:

Hypothesis R5: We expect that there is no significant difference in the perceived experimenter demand effect between treatments. We will test this hypothesis by means of a Kruskal-Wallis test for our three treatments.

Our replication further includes variables 12 [revenge], 13 [punish] and 14 [revenge_others]. These are adapted from Falk et al. (2018, QJE). We include these variables to reveal possible mechanisms for why individuals who dislike or mistrust the EU are reporting higher tail tosses in the EU_Flag treatment (and are therefore more likely to lie in this treatment). We formulate:

Hypothesis R6a: There is a positive correlation between the number of reported tail tosses in EU_Flag and variable 12 [revenge]. That is, the more strongly a subject is a 'revenge type', the greater will be her/his number of reported tail tosses in EU_Flag.

Hypothesis R6b: There is a relatively less strong positive (or no significant) correlation between the number of reported tail tosses in Baseline and Funding and the magnitude of being a 'revenge type'.

Hypothesis R7a: There is a positive correlation between the number of reported tail tosses in EU_Flag and variable 13 [punish]. That is, the more strongly a subject is a 'punishing type', the greater will be her/his number of reported tail tosses in EU_Flag.

Hypothesis R7b: There is a relatively less strong positive (or no significant) correlation between the number of reported tail tosses in Baseline and Funding and the magnitude of being a 'punishing type'.

Hypothesis R8a: There is a positive correlation between the number of reported tail tosses in EU_Flag and variable 14 [revenge_others]. That is, the more strongly a subject is a 'revenge for others type', the greater will be her/his number of reported tail tosses in EU_Flag.

Hypothesis R8b: There is a relatively less strong positive (or no significant) correlation between the number of reported tail tosses in Baseline and Funding and the magnitude of being a 'revenge for others type'.

General remarks

We expect that (almost all) participants should have voted LEAVE in the Brexit referendum. If this is not the case for certain participants, we would also restrict the sample to those who have voted and would vote again for some form of Leave (hard Brexit or May's Leave deal).

In addition to the variables in Table 1, which subjects will report, we will also collect data on the time that subjects spend on the screens. We expect to find stronger results for subjects who do not simply click through the screens as fast as possible and who take the study seriously. If we find an indication of fast clicking, we may exclude those subjects as a robustness check.