Volume 38, Issue 3

Video recordings in experiments – Are there effects on self-selection or the outcome of the experiment?

Tim Lohse
Berlin School of Economics and Law

Salmai Qari
Berlin School of Economics and Law

Abstract

The use of video recordings in experimental economics has become increasingly popular. However, little attention is paid to how this might affect the composition of the participating subjects and the intended treatment effect. We make a first attempt to shed light on these issues and address them in an incentivized face-to-face tax compliance experiment. The experiment contains two dimensions; i) the level of the fine for non-compliance; and ii) the presence of a recording video camera. The 2x2 design frees the intended treatment effect of the fine from any effect resulting from the announced use of a camera. Our findings point in the direction that neither gender nor personality traits nor other individual characteristics seem to have the explanatory power to predict participation in sessions' with or without a camera, respectively. Most importantly, the presence of a recording video camera does not affect subjects' observed decision behavior in the actual experiment.

1. Introduction

Since the groundbreaking experimental work by Ekman and Friesen (1974) on lie catching, the analysis of video (or even TV) recordings of (laboratory) subjects has developed as a promising method in social psychology and, more recently, in experimental economics. Besides studying verbal and nonverbal deception cues, current research uses recordings to address a variety of topics. These include motives for group decision making (Bosman, Hennig-Schmidt, and Van Winden, 2006) and for rejecting advantageous offers in ultimatum games (Hennig-Schmidt, Li, and Yang, 2008), the informativeness of voluntary versus involuntary promises (Belot, Bhaskar, and van de Ven, 2010), the effects of communication on donations to, and discrimination between potential receivers (Greiner, Güth, and Zultan, 2012), the sources of advantage to attractive people (Belot, Bhaskar, and Van De Ven, 2012a), the accuracy of predictions of subjects' trustworthiness and honesty (Belot, Bhaskar, and Van De Ven, 2012b; Konrad, Lohse, and Qari, 2014), the extent to which subjects can successfully delude others (Dwenger and Lohse, 2016) or the measurement of emotions in order to infer subjects' levels of satisfaction and willingness to pay (Gneezy, Gneezy, Lull, and Rey Biel, 2016a,b). This extensive foundation of video recordings literature is accompanied by two methodological questions. First, does prior notification of video recordings during the course of an experiment lead to a self-selection of individuals who sign up to participate in the experiment? Second, does the participants' decision behavior and, therefore, the outcome of an experiment depend on whether or not there is a video camera which records the subjects? These are important questions for the both internal and external validity of laboratory experiments. Structural differences in the set of participating subjects in experiments with video recordings, as compared to experiments without recordings, may falsify the actual treatment effects and, consequently, provide misleading evidence. The same problems arise if subjects' experimental decisions are influenced, or even determined by, the mere presence of a camera recording their behavior. As videotaping may be a technical necessity in many fields of research, it should hardly have any impact on the experiment itself - a widely accepted requirement that somewhat lacks clear evidence (Potter, 1996). Despite the popularity of using video recordings the literature remains surprisingly silent on the aforementioned two questions.[1]

This experimental paper makes a first attempt to shed light on these issues. We use a 2x2 design of an incentivized face-to-face tax compliance game. Subjects have to report their income while facing a potential audit which would detect underreporting. In one dimension, we generate a treatment effect that has been well studied in the tax

---

[1]Only a few recent papers have investigated somewhat related aspects. Belot, Duch, and Miller (2015) and Choo, Fonseca, and Myles (2016) show how different subject pools may lead to different outcomes in the same experimental game. Witt, Donnellan, and Orlando (2011) and Ebersole, Atherton, Belanger, Skulborstad, Allen, Banks, Baranski, Bernstein, Bonfiglio, Boucher, et al. (2016) study gender differences and differences in the personality traits of subjects in an experimental subject pool to predict who prefers in-person participation rather than online participation and who prefers early rather than late participation in the semester, respectively. Bischoff and Frank (2011) check for distortions of the treatment effect through instructors' nonverbal cues but do not find any evidence of this.

compliance literature. We apply two levels of a fine for underreporting and expect less underreporting for the higher fine. The innovative contribution of this paper lies in the second dimension, which is given by the presence or absence of a camera. This camera records the face-to-face income reporting. Contrasting the characteristics of participants who freely signed up for sessions with video recording with those who chose sessions without recording, allows us to gain some insights into potential self-selection effects among various dimensions.[2] Most importantly, comparing reporting behavior for both levels of the fine when a camera is present and when it is not present, respectively, allows us to determine whether the video recording itself affects and, thereby, distorts the intended treatment effect of the experiment, i.e., changes in reporting behavior given the level of the fine.

2. Methodology

2.1. The experimental set-up

We ran an incentivized tax compliance experiment, in which a 'taxpayer' reported her income face-to-face to a 'tax inspector'. A session consisted of four rounds. At the beginning of each round, via a personal monitor, taxpayers were informed about their individual, randomly assigned laboratory income, which was either high (1,000 experimental currency units (ECU) = EUR 16, 80 percent chance) or low (ECU 400 = EUR 6.4, 20 percent chance). The taxpayers' actual income was private information. Tax inspectors only knew the chances of a high or low income, respectively. Taxpayers were called up via their monitor. They were then asked to report an income individually to a tax inspector in a separate room. Each report took about 20 seconds. In each of the four rounds of the experimental session, taxpayers met a different tax inspector. Taxpayers with a low income had a unique best choice; reporting a low income which resulted in a zero statutory tax liability.[3] A taxpayer with a high income could truthfully report the high income and pay a statutory tax of ECU 200 (= EUR 3.2). Alternatively, she could claim to have a low income and thereby try to evade taxes. At the end of each round, once all taxpayers had declared their income, a partial audit via the computer system took place. Taxpayers knew ex ante that tax inspectors influenced the choice of whom to select for an audit: inspectors had to assess a series of up to 10 declarations and grade taxpayers with respect to perceived truthfulness (grades ranging from 1 = very truthful to 10 = very untruthful). Then, in a first step, the computer system identified taxpayers with a high income but reported a low one. This identification was not communicated to anyone (neither to the inspectors nor to the taxpayer herself or other participants) and was a pure technical necessity for the second step. In that second step, half of the underreporting taxpayers were selected for an audit. This selection occurred according

---

[2]Our experiment is just an initial step in studying the self-selection problem. To get a complete picture one would have to analyze the behavior and characteristics of subjects who would have been randomly assigned in sessions with and without recordings after they had signed up.

[3]In the actual experiment there were only three out of 917 observations in which low-income taxpayers did not behave in line with this dominant strategy.

to the ranking made by the tax inspector: the computer system selected those 50 percent of underreporting taxpayers for an audit who were perceived by the tax inspector as being the most untruthful. High-income taxpayers who considered underreporting thus had to take into account how their appearance and cheating performance would affect their individual audit probability, knowing that half of all cheaters would get detected. Taxpayers caught underreporting had to pay the tax of ECU 200, topped up by a fine. The fine for underreporting (if caught) varied: in half of all sessions it was low (ECU 100 = EUR 1.6), otherwise it was high (ECU 300 = EUR 4.8).

The experiment followed a 2x2 between-subjects design where the variation of the fine represents the first dimension. The main innovation of the paper emerges in the second dimension. In some sessions a camera was placed beside the tax inspector recording the taxpayers' reports. In the first invitation email, sent to the entire subject pool in order to advertise the experiment, we announced that video recordings would take place in some sessions. This ex ante information was required by the privacy policy of the university. Thus, when signing up for the experiment subjects could freely choose whether to enroll for a session with video recording or without video recording. They were informed that earning schemes would not depend on the recordings. Before the actual experiment started, subjects participating in the treatments with a camera were asked to sign a declaration of consent that a recording would be made in the experiment and that these video clips could be used by the experimenter for further scientific purposes.[4] Furthermore, we informed them that they could revoke their consent and leave the experiment at any time during the experiment. In such a case they would only receive the show-up fee. Despite this, no participants left and everyone that opted into being videotaped agreed to the recording.

After the experiment, the subjects had to answer two additional questionnaires containing socio-economic questions and questions of the "BIG 5" approach, a psychological concept used to describe and study personality (McCrae and John, 1992; Borghans, Duckworth, Heckman, and Ter Weel, 2008). The BIG 5 approach relies on the assumption that personality differences between individuals, which are manifested in different ways of behaving and experiencing the world, are due to differences among five basic personality traits: openness to experience, conscientiousness, extraversion, agreeableness. and neuroticism. We use a short 15-item version of common inventories as is also applied in the German Socio-Economic Panel (SOEP) (Gerlitz and Schupp, 2005; Dehne and Schupp, 2007). We apply principal component analysis to extract the underlying five personality factors for each subject.[5] The questionnaires were followed by a standard risk elicitation game in the style of Holt and Laury (2002). The generated measure of risk aversion is simply the number of safe choices for each subject.

---

[4]The conducted experiment is part of a larger research project on compliance (Konrad et al., 2014; Konrad, Lohse, and Qari, 2017). Konrad et al. (2014) study the self-selection of capable and less capable deceivers whereas Konrad et al. (2017) focus on the implications of endogenizing the audit probabilities in a tax compliance game. The face-to-face setting analyzed in this paper resembles treatment T3 in Konrad et al. (2017).

[5]We use the R environment (R Core Team, 2017) to conduct the analyses. The principal component analysis was performed with the package by Revelle (2017).

The experiment took place in several waves at the MELESSA, the experimental laboratory at the University of Munich. Sessions of the respective treatments were scheduled on different days and at different times of the day. A total number of 295 subjects participated in the role of taxpayers, of which 98 subjects are in "low fine/no camera", 68 in "low fine/camera", 79 in "high fine/no camera" and 50 in "high fine/camera". Each session had up to 20 taxpayers who were students from various fields of study recruited by the MELESSA laboratory using the software ORSEE (Greiner, 2015). Tax inspectors were student assistants of the Max Planck Institute for Tax Law and Public Finance in Munich. Participants acting as taxpayers received their earnings from the experiment (a single round was randomly selected to determine subjects' payoffs), plus an average income from performing the risk elicitation game of EUR 2, plus a show-up fee of EUR 4. They earned EUR 18.06 on average with a standard deviation of EUR 3.46.[6]

2.2. Hypotheses

To address our research questions we developed two hypotheses. First, since individuals could enroll for experimental sessions with a video recording or without, we expect some self-selection of the participants. Different subsets of the subject pool might participate in the two kinds of sessions. On the one hand, some individuals might e.g., be curious about the novel approach of using cameras in the lab, or simply like the idea of being in front of a camera. These individuals could be more likely to select sessions with the video recordings. In terms of the BIG5 personality traits they might reveal a higher degree of openness to experience and extraversion. On the other hand, some individuals might be opposed to leaving some more permanent traces of their participation in the experiment, or may simply dislike the idea of being recorded. Individuals of this type would prefer sessions without video recordings and might show e.g., a higher degree of neuroticism and risk aversion. Potentially, preferences for settings with or without a camera might be more pronounced with one or the other gender, or could even depend on personal characteristics such as age or religiosity. Taken together, we state the following hypothesis:

Hypothesis 1 (Ex ante camera effect on self-selection): *The sets of subjects participating in sessions with and without video recordings, respectively, may differ with respect to the gender composition, the subjects' personality traits or personal characteristics.*

Second, to see if and how a recording camera affects subjects' decision behavior in an experiment, we chose a tax compliance framework. Understanding such a framework is intuitive and therefore rather easy. Of the two dimensions of our experiment, the variation of the fine is standard in (tax) compliance literature. It has been found to impact subjects' behavior in an unambiguous way: a higher fine for underreporting results in

---

[6]In line with reality, the four tax inspectors serving per session were paid a flat rate. A further incentivation of the inspectors was not necessary since the research focus was entirely on the reporting individuals. Besides, recall that each tax inspector encountered each taxpayer just once.

comparably more honest income reporting (Alm, Jackson, and McKee, 1992; Torgler, 2002). For our study, this intended treatment effect only serves as a control when varying the second dimension. Besides the fine, the mere presence of a recording video camera may affect subjects' compliance decisions, too. Although subjects knew that whoever saw their videotape would remain uninformed about whether their statement was true or false, a deceptive strategy could come at higher psychological costs. Reputational concerns or shame could play a larger role than in sessions without a camera and, hence, lead to more honest behavior. Furthermore, these psychological costs could be larger if the fine is high because a higher fine indicates a behavior that is even more socially condemned. We summarize these considerations as follows:

Hypothesis 2 (Camera effect on decision behavior): *Deceptive behavior occurs less often when a video camera records subjects' statements. This effect is exacerbated when the fine is high.*

In sum, if there is evidence for hypothesis 1 or 2 (or even both), the use of a camera in experiments may involve problems of self-selection and treatment bias and, therefore, does not produce reliable insights compared to settings without video recordings.

3. Results

3.1. Selection into sessions

To test hypothesis 1 regarding the possible self-selection of participants, we run a logit model[7] where the dependent variable c_i is a dummy indicating whether subject i participated in sessions in which a camera was present. Formally, the regression equation reads

$$c_i = \alpha_0 + \mathbf{x_i'}\boldsymbol{\delta} + u_i. \tag{1}$$

The explanatory variables collected in $\mathbf{x_i'}$ are a dummy indicating whether the subject is female, a measure for risk aversion, age dummies, indicator variables for religious denomination and the five extracted personality variables. Table 1 presents summary statistics of these variables. If subjects self-select into sessions in which a camera was present, at least one δ_j must be non-zero, i.e., we evaluate hypothesis 1 by testing whether or not the variables collected in $\mathbf{x_i'}$ are jointly significant. Note that we expect some correlation between the explanatory variables. For example, the risk measure is presumably correlated with some personality indicators. Such a correlation would affect individual point estimates and the associated standard errors, but the joint test is not affected.

---

[7]Robust standard errors are implemented via the sandwich package (Zeileis, 2004, 2006).

Table 1: Summary statistics

Variable	Share / Mean	St. Dev.	N
Female	0.593	0.492	295
RiskMeasure	6.597	2.053	295
Age ≤ 20	0.166	0.372	295
Age $\in [21, 23]$	0.427	0.496	295
Age $\in [24, 27]$	0.302	0.460	295
Age $\in [28, 31]$	0.068	0.252	295
Age ≥ 32	0.037	0.190	295
Catholic	0.444	0.497	295
Protestant	0.200	0.401	295
Other Christian	0.044	0.206	295
Muslim	0.007	0.082	295
Other religion	0.014	0.116	295
No religion	0.292	0.455	295
Extraversion	0.0	1.0	295
Conscientiousness	0.0	1.0	295
Neuroticism	0.0	1.0	295
Agreeableness	0.0	1.0	295
Openness to experience	0.0	1.0	295

Notes: The table presents shares (for dummy variables) or means and the associated standard deviations for the explanatory variables. The last listed five personality variables are obtained by running a principal component analysis. They are normalized to a mean of zero and a standard deviation of one.

Table 2: Selection into sessions

	Dependent variable:		
	Dummy variable indicating whether or not a camera was present		
	(1)	(2)	(3)
Female	0.286	0.138	0.161
	(0.262)	(0.298)	(0.315)
RiskMeasure	−0.001	0.002	−0.001
	(0.064)	(0.065)	(0.067)
Age ∈ [21, 23]	−0.425	−0.483	−0.478
	(0.368)	(0.374)	(0.382)
Age ∈ [24, 27]	−0.088	−0.128	−0.099
	(0.385)	(0.394)	(0.405)
Age ∈ [28, 31]	0.196	−0.005	0.005
	(0.580)	(0.591)	(0.602)
Age ≥ 32	−0.298	−0.231	−0.145
	(0.767)	(0.791)	(0.824)
Protestant	0.210	0.151	0.177
	(0.331)	(0.340)	(0.348)
Other Christian	0.196	0.177	0.142
	(0.659)	(0.634)	(0.682)
Muslim	0.479	0.318	0.450
	(2.830)	(2.922)	(2.976)
Other religion	0.454	0.543	0.468
	(1.301)	(1.350)	(1.373)
No religion	−0.013	0.107	0.145
	(0.309)	(0.326)	(0.341)
Extraversion		0.202	0.151
		(0.136)	(0.225)
Conscientiousness		0.268	0.350
		(0.146)	(0.224)
Neuroticism		0.085	0.076
		(0.146)	(0.269)
Agreeableness		0.157	−0.066
		(0.129)	(0.239)
Openness to experience		−0.197	−0.231
		(0.141)	(0.248)
Female*Extraversion			0.096
			(0.289)
Female*Conscientiousness			−0.134
			(0.300)
Female*Neuroticism			0.011
			(0.326)
Female*Agreeableness			0.352
			(0.289)
Female*Openness to experience			0.074
			(0.307)
Constant	−0.423	−0.348	−0.392
	(0.544)	(0.557)	(0.576)
N (# of subjects)	295	295	295
Log Likelihood	-195.966	-190.507	-189.347
Akaike Inf. Crit.	415.933	415.014	422.695
F	0.4	0.76	0.66
Pr(>F)	0.95	0.73	0.87

Notes:

standard errors in parentheses
*p<0.05; **p<0.01; ***p<0.001
Logit model that (does not) predict the probability to select into the camera treatment

The table shows the results from a logit model where the dependent variable is a dummy variable indicating whether or not the subject participated in treatments where a camera was present. The models test whether demographics and personality variables are correlated with the decision to participate in camera-treatments. In all models, the small F-statistics show that the null hypothesis of no correlation is not rejected.

Table 2 compiles the results of the logit model.[8] The first regression (column 1) omits the personality variables and indicates that all variables are individually and in particular jointly insignificant ($F = 0.40$, p-value 0.95). Column (2) includes the five personality variables. The positive extraversion-coefficient may be in line with the idea –mentioned at the beginning of section 2.2.– that extraversion and camera-session-participation are correlated. However, as before, all variables are jointly insignificant ($F = 0.76$, p-value 0.73). Finally, column (3) considers interaction terms between the female-dummy and the personality variables. This generates the same finding of neither individual nor joint predictive power ($F = 0.66$, p-value 0.87). In summary, in all three specifications the null hypothesis of no correlation between the observables and the treatment indicator is not rejected. Therefore, in our experimental setting we do not find evidence for hypothesis 1. Obviously, it is still possible that there may be self-selection for unobservable characteristics. Moreover, as mentioned above, a more sophisticated experimental setup might provide a more complete picture.

3.2. Outcome of the experiment

In order to evaluate hypothesis 2, we analyze the compliance decisions of the taxpayers. As explained in the experimental setup, each taxpayer played four rounds and randomly received either a low or a high income in each round. If the taxpayer receives a high income, she may underreport. Therefore, for the following analysis we select all rounds in which subjects had a high income. Since this occured with an 80 percent chance, the *expected* number of observations is given by 295·4·0.8. The *actual* number of observations in our sample is equal to 917. Table 3 provides the aggregate rate of honest reporting

Table 3: Aggregate compliance

	(1) No Camera	(2) Camera
Low Fine	0.43	0.36
High Fine	0.74	0.67

in the four treatments. For example, when the fine is low and there is no camera in the room, the fraction of honest reporting decisions is equal to 43 percent. In contrast, if the fine is high, this aggregate honest reporting is 31 percentage points larger and equal to 74 percent. If the fine is low and there is a camera present, aggregate compliance is equal to 36 percent. This compliance increases by 31 percentage points and is equal to 67 percent if the fine is high. These descriptive results do not indicate that the camera leads to a treatment bias; the intended treatment effect is the change in compliance when the fine is high rather than low. Whether or not a camera is present, aggregate compliance is 31 percentage points higher if the fine is high rather than low. Since the treatment bias is measured by the difference-in-differences, the descriptives suggest a treatment bias of zero. In order to test whether any of these differences are statistically

---

[8]The tables were generated with the stargazer package (Hlavac, 2015).

significant, we ran a series of logit models. The dependent variable is a dummy variable indicating compliance, i.e, the dummy is equal to 1 if the taxpayer honestly reports his or her true high income and equal to zero if the taxpayer underreports. The primary explanatory variables are three dummy variables according to the 2x2 cases shown in table 3. Formally, the regression equation reads

$$y_{it} = \beta_0 + \beta_1 * \text{High Fine} + \beta_2 * \text{Camera} + \beta_3 * \text{Camera*High Fine} + \mathbf{\Xi'_{it}}\boldsymbol{\gamma} + u_i + \epsilon_{it}, \quad (2)$$

where $\mathbf{\Xi'_{it}}$ comprises additional control variables like gender and age. u_i denotes the subject-specific random effects that take into account the repeated measurement of the same subject. In this specification, a significant β_3 would indicate that the camera-effect in the high-fine treatment differs from the camera-effect in the low-fine treatment, while β_2 models a possible common camera-effect for both the low- and high-fine-treatments.

Table 4 compiles the results. In the first column, we enter only two dummies indicating the high fine and the presence of the camera, while the second column additionally includes the interaction term between Camera and High Fine. These two models indicate that the camera-slopes are not significantly different from zero. Column (2) additionally shows that the interaction term is insignificant. Therefore, column (2) complements the descriptives in table 3 and shows that compliance behavior is not significantly affected by the presence of the camera, i.e., the compliance rate of 43 percent is statistically not different from 36 percent, and 74 percent is statistically not different from 67 percent.

The high-fine coefficient is similar across the specifications in the first two columns and translates into an economically large treatment effect. For example, in column (1), the constant term (-0.611) indicates that the average rate of honest reporting is equal to 42 percent if the fine is low.[9] The high-fine coefficient translates into an average honest reporting of 82 percent. Hence, the intended treatment effect of the experimental setup is equal to 40 percentage points. This effect is precisely estimated and significant at conventional levels. This sizable effect resembles the descriptives (table 3).

The third column probes for possible gender effects of the taxpayer and the tax inspector. It further tests possible tax reporting dynamics by entering a dummy indicating whether or not the tax reporting takes place in the second half of the experiment. The associated Akaike information criteria indicate that these three additional variables do not jointly provide additional explanatory power, as the model in column (2) is preferred compared to column (3). In columns (4) and (5) we enter further control variables like the measure capturing risk aversion, age dummies, and the personality variables. Column (4) omits the two reporting-specific variables, while column (5) considers all variables. The two columns indicate that a higher degree of risk aversion increases compliance and they show a correlation between age and compliance. Once again, the associated AIC indicate that the model in column (4) without the reporting-specific variables is preferred. More importantly, the two columns provide the same qualitative findings regarding the main treatment variables.

---

[9]We obtain this average by simulating 500,000 draws from the distribution of the subject-specific random effects; this allows us to calculate 500,000 individual-specific reporting probabilities and to calculate the corresponding average across the 500,000 individuals.

Table 4: Outcome of the experiment

	Dependent variable:				
	Dummy variable indicating whether or not a subject honestly reports a high income				
	(1)	(2)	(3)	(4)	(5)
High Fine	2.635***	2.815***	2.867***	2.404***	2.423***
	(0.415)	(0.526)	(0.537)	(0.526)	(0.532)
Camera	−0.568	−0.390	−0.305	−0.555	−0.461
	(0.367)	(0.479)	(0.489)	(0.504)	(0.513)
High Fine * Camera		−0.430	−0.501	−0.041	−0.079
		(0.746)	(0.757)	(0.785)	(0.794)
laterRounds			−0.227		−0.267
			(0.233)		(0.238)
maleOfficer			0.170		0.177
			(0.218)		(0.223)
Female			0.210	0.138	0.150
			(0.375)	(0.427)	(0.432)
RiskMeasure				0.210*	0.213*
				(0.094)	(0.095)
Age $\in [21, 23]$				0.955	0.970
				(0.556)	(0.562)
Age $\in [24, 27]$				1.809**	1.838**
				(0.600)	(0.607)
Age $\in [28, 31]$				1.922*	1.962*
				(0.897)	(0.907)
Age ≥ 32				1.613	1.644
				(1.059)	(1.070)
Extraversion				0.112	0.109
				(0.189)	(0.191)
Conscientiousness				−0.132	−0.130
				(0.203)	(0.205)
Neuroticism				0.169	0.168
				(0.201)	(0.203)
Agreeableness				−0.095	−0.094
				(0.188)	(0.189)
Openness to experience				−0.028	−0.028
				(0.187)	(0.189)
Constant	−0.611*	−0.688*	−0.835*	−3.133***	−3.185***
	(0.286)	(0.317)	(0.407)	(0.869)	(0.883)
Observations	917	917	917	917	917
Log Likelihood	−502.142	−501.975	−501.053	−484.555	−483.607
Akaike Inf. Crit.	1,012.285	1,013.950	1,018.106	1,001.110	1,003.213
Bayesian Inf. Crit.	1,031.569	1,038.055	1,056.675	1,078.247	1,089.993

Notes: standard errors in parentheses
*p<0.05; **p<0.01; ***p<0.001
Logistic mixed effects model with subject-specific random effects

The table shows the results from a logit model where the dependent variable is a dummy variable indicating whether a subject with a high income truthfully declares this income. The two main models in columns (1) and (2) indicate no significant changes in the compliance behavior induced by the camera. The model in column (3) further indicates a non-significant interaction term, and therefore there is no evidence that the placement of a camera affects compliance behavior differently in the low-fine and the high-fine treatments. The remaining models test the robustness of these findings by including demographics like gender and personality variables. The camera-slope and the corresponding interaction term are not significant in these robustness specifications.

In summary, in all models the slopes β_2 and β_3 are not statistically different from zero and, therefore, there is no evidence for hypothesis 2. This finding is unaffected by including control variables like age, gender or personality variables.

4. Conclusion

This paper investigates two questions which arise in economic experiments that make use of the video recordings of participants: first, whether prior notification of video recordings in the course of an experiment leads to a self-selection of individuals who sign up for the experiment. Second, if the participants' decision behavior and, therefore, the outcome of an experiment depend on whether or not subjects are videotaped. In a face-to-face income reporting experiment, in the style of a tax compliance game, we vary two dimensions: the level of the fine for underreporting and the presence of a recording video camera. This 2x2 design allows for the disentangling of the intended treatment effect of the fine, from effects resulting from the announced use of a camera. Our results do not indicate self-selection into sessions with or without a camera, respectively. We do not find correlation between personality traits, or other observable characteristics, and participation in experiments involving the camera. Therefore, at least in our experimental setting, the subjects are not systematically different when videotaping takes place. Most importantly, the presence of a recording video camera does not interfere with subjects' decision behavior in the actual experiment. In the case of a low fine, subjects underreport more, and in the case of a high fine they underreport less - irrespective of whether their actions were caught on camera or not.

In evaluating our findings, some caveats come to mind. First, our analysis is only a first step in studying the self-selection problem. A definite conclusion would require running an additional experimental setting in which subjects who signed up for the experiment would be randomly assigned in sessions with or without cameras. Investigating their behavior and their characteristics would give a complete picture of whether self-selection occurs or not. Second, we chose a tax compliance framework, as it is well studied and intuitive for the participants. This is a specific experimental setup and the questions addressed in this paper could be studied in different contexts as well. Third, we only focused on the most important socio-economic characteristics and five personality traits, as captured by the BIG5. Future research might take into account many more aspects. Lastly, our findings speak to researchers who use cameras to record subjects so as to analyze the recordings afterward. A related strand of the experimental literature uses cameras only to allow instant communication among participants via video conferencing (Brosig, Weimann, and Ockenfels, 2003; Schmidt and Zultan, 2005; Zultan, 2012). The problem of a potential selection effect could arise here, too. Our results provide reason to believe that there should not be any selection effects even for this literature. However, due to the different experimental setting it is not definitely clear that findings carry over.

References

Alm, J., B. R. Jackson, and M. McKee (1992). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal* **45**, 107–114.

Belot, M., V. Bhaskar, and J. van de Ven (2010). Promises and cooperation: Evidence from a tv game show. *Journal of Economic Behavior & Organization* **73**(**3**), 396–405.

Belot, M., V. Bhaskar, and J. Van De Ven (2012a). Beauty and the sources of discrimination. *Journal of Human Resources* **47**(**3**), 851–872.

Belot, M., V. Bhaskar, and J. Van De Ven (2012b). Can observers predict trustworthiness? *Review of Economics and Statistics* **94**(**1**), 246–259.

Belot, M., R. Duch, and L. Miller (2015). A comprehensive comparison of students and non-students in classic experimental games. *Journal of Economic Behavior & Organization* **113**, 26–33.

Bischoff, I. and B. Frank (2011). Good news for experimenters: Subjects are hard to influence by instructors' cues. *Economics Bulletin* **31**(**4**), 3221–3225.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources* **43**(**4**), 972–1059.

Bosman, R., H. Hennig-Schmidt, and F. Van Winden (2006). Exploring group decision making in a power-to-take experiment. *Experimental Economics* **9**(**1**), 35–51.

Brosig, J., J. Weimann, and A. Ockenfels (2003). The effect of communication media on cooperation. *German Economic Review* **4**(**2**), 217–241.

Choo, C. L., M. A. Fonseca, and G. D. Myles (2016). Do students behave like real taxpayers in the lab? evidence from a real effort tax compliance experiment. *Journal of Economic Behavior & Organization* **124**, 102–114.

Dehne, M. and J. Schupp (2007). Persönlichkeitsmerkmale im Sozio-oekonomischen Panel (soep)-Konzept, Umsetzung und empirische eigenschaften. *Research Notes* **26**, 1–70.

Dwenger, N. and T. Lohse (2016). Do individuals put effort into lying? Evidence from a compliance experiment. *CESifo Working Paper No.* **5805**.

Ebersole, C. R., O. E. Atherton, A. L. Belanger, H. M. Skulborstad, J. M. Allen, J. B. Banks, E. Baranski, M. J. Bernstein, D. B. Bonfiglio, L. Boucher, et al. (2016). Many labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology* **67**, 68–82.

Ekman, P. and W. V. Friesen (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology* **29**(**3**), 288.

Gerlitz, J.-Y. and J. Schupp (2005). Zur erhebung der big-five-basierten persoenlichkeitsmerkmale im soep. *DIW Research Notes* **4**, 2005.

Gneezy, U., A. Gneezy, J. Lull, and P. Rey Biel (2016a). Expectations, satisfaction and utility from experience goods: A field experiment in theaters. *Barcelona GSE Working Paper No.* **944**.

Gneezy, U., A. Gneezy, J. Lull, and P. Rey Biel (2016b). Utility and willingness-to-pay through emotion measuring and pay-what-you-want pricing.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* **1**(**1**), 114–125.

Greiner, B., W. Güth, and R. Zultan (2012). Social communication and discrimination: a video experiment. *Experimental Economics* **15**(**3**), 398–417.

Hennig-Schmidt, H., Z.-Y. Li, and C. Yang (2008). Why people reject advantageous offers—non-monotonic strategies in ultimatum bargaining: Evaluating a video experiment run in pr china. *Journal of Economic Behavior & Organization* **65**(**2**), 373–384.

Hlavac, M. (2015). *stargazer: Well-Formatted Regression and Summary Statistics Tables.* Cambridge, USA: Harvard University. R package version 5.2.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review* **92**(**5**), 1644–1655.

Konrad, K. A., T. Lohse, and S. Qari (2014). Deception choice and self-selection - the importance of being earnest. *Journal of Economic Behavior & Organization* **107A**, 25–39.

Konrad, K. A., T. Lohse, and S. Qari (2017). Compliance with endogenous audit probabilities. *The Scandinavian Journal of Economics* **119**, 821–850.

McCrae, R. R. and O. P. John (1992). An introduction to the five-factor model and its applications. *Journal of Personality* **60**(**2**), 175–215.

Potter, J. (1996). Discourse analysis and constructionist approaches: Theoretical background. British Psychological Society.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2017). *psych: Procedures for Psychological, Psychometric, and Personality Research.* Evanston, Illinois: Northwestern University. R package version 1.7.8.

Schmidt, C. and R. Zultan (2005). The uncontrolled social utility hypothesis revisited. *Economics Bulletin* **3**, 1–7.

Torgler, B. (2002). Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys* **16**(**5**), 657–683.

Witt, E. A., M. B. Donnellan, and M. J. Orlando (2011). Timing and selection effects within a psychology subject pool: Personality and sex matter. *Personality and Individual Differences* **50**(**3**), 355–359.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software* **11**(**10**), 1–17.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software* **16**(**9**), 1–16.

Zultan, R. (2012). Strategic and social pre-play communication in the ultimatum game. *Journal of Economic Psychology* **33**(**3**), 425–434.