



Working Papers

www.cesifo.org/wp

Dubious Versus Trustworthy Faces – What Difference Does it Make for Tax Compliance?

Kai A. Konrad
Tim Lohse
Salmai Qari

CESIFO WORKING PAPER NO. 4373

CATEGORY 1: PUBLIC FINANCE

AUGUST 2013

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

CESifo

Center for Economic Studies & Ifo Institute

Dubious Versus Trustworthy Faces – What Difference Does it Make for Tax Compliance?

Abstract

We find experimental evidence that the decision problem of tax compliance changes if subjects' declarations are not randomly assessed, but is based on their appearance as captured by pictures of their faces, even if the aggregate audit probability does not change. Some subjects may fear that their picture looks rather dubious whereas others may believe that their picture looks more trustworthy than average. Depending on these beliefs, they may adjust their compliance decisions. Our experimental design allows to disentangle these potentially countervailing effects.

JEL-Code: H260, C910, H310.

Keywords: compliance, tax evasion, face picture, audit probability.

Kai A. Konrad

Max Planck Institute for Tax Law and Public Finance

Marshallplatz 1

Germany – 80539 Munich

kai.konrad@tax.mpg.de

Tim Lohse

Berlin School of Economics and Law

Alt-Friedrichsfelde 60

Germany – 10315 Berlin

tim.lohse@hwr-berlin.de

Salmai Qari

Max Planck Institute for Tax Law and

Public Finance

Marshallplatz 1

Germany – 80539 Munich

salmai.qari@tax.mpg.de

August 21, 2013

For providing laboratory resources we kindly thank MELESSA of the University of Munich. We thank Hans Müller for developing and programming the web-based environment. We thank Sophia Baur, Nina Bonge, Ludwig Grill, David Houser, Sarah Marfeld, Daniela Miehl, Christoph Rüschoer, Verena Schönecker, Maria Selmsberger, and Jenny Zeiser for excellent research assistance. We also thank participants of the APET meeting 2013 in Lisbon for helpful comments. The usual caveat applies.

1 Introduction

When people present themselves on social networks such as Facebook or LinkedIn, a photo of a person's face is typically among the first crucial pieces of information. A picture is worth a thousand words and persons' looks affect economic behavior of their counterparts.¹ In trust games subjects are willing to pay a fee to see the pictures of their partners and this option affects the outcome (Eckel and Petrie 2011). We consider a related question in compliance games: Does it affect a person's tax compliance behavior if the person knows that the tax officer can see the person's picture? Whether or not their picture is seen by the tax inspector changes the person's probability beliefs about being audited if she underreports, suggesting that it affects her compliance behavior. If a subject's picture is seen by the tax officer, and if it affects the tax officers assessment, the direction of the effect is unclear. Some subjects may fear that their picture looks more suspicious and dubious than average, others may be convinced that their picture looks more trustworthy than average. Depending on these beliefs, they may adjust their compliance decisions. Our experimental design takes these potentially countervailing reactions into consideration and allows for a separate measurement of the two types of reaction. The main findings are twofold. First, there is evidence for both types of reactions. Second, the effect of subjects who believe that their picture looks more trustworthy than average is stronger than the opposite effect.

2 Experimental design and predictions

Our experimental design borrows from the compliance set-up in Konrad, Lohse and Qari (2012). Subjects are assigned an endowment randomly. They learn whether their own endowment of some taxable good is high (1000 Taler with an 80% chance) or low (400 Taler with a 20% chance). Every subject must then declare her own endowment as either high or low. Low endowments are tax/duty free. If a subject declares high endowment, a tax/customs duty of 200 Taler is levied.² If a subject reports a low endow-

¹For instance: beauty (Mobius and Rosenblat 2006, Wilson and Eckel 2006), ethnicity (Habyarimana et al. 2007), gender (Solnick and Schweitzer 1999, Andreoni and Petrie 2008), and race (Castillo and Petrie 2010) have been identified as relevant aspects of appearance. DeBruine (2002) reports evidence from a trust game and argues that players trust co-players more who, according to their picture, resemble themselves.

²Despite the complexity of tax evasion the experimental literature by and large reduces the related compliance decision to a question about whether or not to cheat. In this sense, we do not only address customs compliance, but the more general matter of tax evasion.

ment, an audit may follow. An audit perfectly reveals the true endowment, if it takes place. If the subject underreports and receives an audit, the subject is caught and must pay the duty of 200 Taler plus a penalty tax.

A BASE treatment follows the standard approach of tax compliance games: The subjects sit in front of the computer, read their own true endowment on the screen, type in their compliance choice, and the computer randomly selects half of all underreporters for an audit. In this BASE treatment the audit probability (and detection that is implied by an audit) is exogenously determined and executed by the computer: half of all underreporters receives an audit. This is public information.

The innovation of this paper occurs in the target treatment, PIC: The subjects' faces are photographed prior to the actual experiment. A student assistant in the role of a compliance officer sees a series of (up to ten) pictures of subjects on his screen, who reported "low". The assistant knows all these subjects reported "low" - some truthfully, and some of them may have underreported. It is the assistant's task to rank the pictures of this group according to the perceived likelihood that the respective subject underreported. The audit procedure in PIC leads to the situation that exactly one half of all underreporters receive an audit, just as in the BASE treatment, but the selection procedure is different and increases audit probability for subjects ranked as likely underreporters.³

The subjective probability for receiving an audit should matter for the compliance decision. In the BASE treatment, subjects may, but need not think that their subjective probability equals the technologically implemented probability of $1/2$.⁴ Deviations from $1/2$ are more natural in the PIC treatments where the subjects may form subjective beliefs about how their own picture affects their own audit probability. Deviations from $1/2$ may occur in both possible directions. Some subjects may think that their picture makes them more trustworthy in the eyes of the tax inspector, for other subjects the reverse type of beliefs may be formed:

³Coricelli et al. (2010) use images of faces in a compliance experiment, as well, but with a rather different focus: An individual's cheating behavior is revealed by publicly displaying his picture in the laboratory. The risk of public exposure of deception deters evasion significantly.

⁴In the BASE treatment the audit probability for each subject reporting low income is exogenously and technologically fixed at $1/2$, and subjects are informed about this. This does not imply, however, that subjective probability assessments cannot deviate from $1/2$. Some subjects may, for instance, simply believe that Fortuna favors them or does not favor them. However, our analysis focuses on the treatment difference in the share that has beliefs that deviate from $1/2$.

Hypothesis 1A: The share of subjects in the PIC treatments who believe their probability is lower than $1/2$ should be higher than in the BASE treatments.

Hypothesis 1B: The share of subjects in the PIC treatments who believe their probability is higher than $1/2$ should be higher than in the BASE treatments.

In summary, Hypotheses 1A and 1B imply that the share of subjects in the PIC treatments who believe their probability deviates from $1/2$ should be higher than in the BASE treatments.

This probability assessment may also affect compliance decisions. In line with Hypothesis 1A, subjects who consider an audit to be less likely may be more inclined to underreport. And in line with Hypothesis 1B, subjects who consider an audit to be more likely may be less inclined to underreport.

To distinguish between the treatment effect of PIC for subjects who revise their audit probability upward and for subjects who revise it downward, we consider two variants of BASE and PIC which differ in the size of the fine. BASE-low and PIC-low have a low penalty fine (100 Taler), BASE-high and PIC-high have a high penalty fine (300 Taler). The variant with a low fine will be suitable for identifying subjects who revise their audit probability upward, and the high-fine variant will be suitable to identify subjects who revise their audit probability downward. The mechanism by which this works is as follows: In case of a low fine, as in BASE-/PIC-low, if we focus on monetary concerns only, a risk neutral subject who considers the subjective audit probability to be close to $1/2$ would be inclined to underreport. Let us call this choice the "theory prediction" for BASE-low. From the perspective of monetary payoff, in BASE-low a subject would be just indifferent between underreporting and complying if she believes that she will receive an audit with a probability of $2/3$ in case of underreporting and should prefer to underreport for all subjective audit probabilities smaller than $2/3$. In fact, in BASE-low we find a very high share of subjects underreporting, which is in line with subjective probability beliefs smaller than $2/3$, and in particular, with probability beliefs of about $1/2$. In the PIC treatment, in line with Hypothesis 1A, there may be a share of players who attribute a subjective audit probability to them that is considerably higher than $1/2$. This group of players may deviate from the theory prediction. Accordingly, in the PIC-low treatment subjects who have a sufficiently higher subjective audit probability in line with Hypothesis 1A would have underreported in BASE-low but report truthfully in PIC-low. The comparison between BASE-/PIC-low is therefore suitable to pick up the behavioral change that occurs

for subjects who strongly increase their subjective probability for receiving an audit once their picture is shown to a tax inspector.

In case of a high fine, as in BASE-/PIC-high, for analogous reasons the data can pick up the behavioral change of the group of subjects who think their picture, if shown to the tax officer, will downgrade their audit probability. Here, in the BASE-treatment subjects who think that their audit probability is equal or close to $1/2$ and who primarily care for monetary rewards would report truthfully. Only subjects who have a subjective probability for an audit below $1/3$ may report truthfully in BASE-high but underreport in PIC-high. Such strong deviations in the subjective probabilities from what is the average, laboratory-induced audit probability may occur more frequently in the PIC treatments in line with Hypothesis 1B. Accordingly, the comparison between BASE-high and PIC-high is suitable to identify this type of behavioral variation.

Corresponding to Hypotheses 1A and 1B, overall, we expect that PIC compared to BASE has a larger spread in the subjective probability estimates. These, in turn, should induce an increase in truthful compliance in PIC-low compared to BASE-low, and a reduction in truthful compliance in PIC-high compared to BASE-high:

Hypothesis 2A: The rate of truthful reporting among high-income subjects is higher in PIC-low than in BASE-low.

Hypothesis 2B: The rate of truthful reporting among high-income subjects is lower in PIC-high than in BASE-high.

In total 144 subjects participated in the low-penalty setup and 94 subjects participated in high-penalty setup. They were recruited⁵ from the general student subject pool of the MELESSA laboratory in Munich, where we also conducted the experiment. Each subject participated in one treatment only. The BASE treatment was played for 10 rounds, the PIC treatment for 4 rounds.⁶ This led to 649 compliance decisions for subjects with high income in low-penalty set-ups and 499 compliance decisions for subjects with high income in high-penalty set-ups. In each round of the PIC treatment, subjects are ranked by a different student assistant, to rule out strategic interaction in between rounds. Subjects filled in a questionnaire after completing all compliance treatment rounds and participated in a standard risk

⁵The participants were recruited using the software ORSEE (Greiner 2004).

⁶Since the BASE treatment takes less time than the target treatment, PIC, subjects participate in ten rounds in the BASE treatment, but they participate in only four rounds in treatment PIC.

elicitation game in the style of Holt and Laury (2002). A subject's final payoff consisted of the income of a single (randomly chosen) round, net of taxes and net of fines in case of an audit, plus the income from participating in the risk elicitation game and a show-up fee.

3 Results and discussion

We first analyze the difference in subjects own audit probability assessments for the BASE treatments and the PIC treatments. The relevant question asks subjects to "assess their own subjective audit probability in case they consider to underreport"; subjects answer by selecting one of the three categories "lower than 1/2", "equal to 1/2" and "higher than 1/2". We code the possible answers as -1 for an audit probability lower than 1/2, 0 for an audit probability equal to 1/2 and 1 for an audit probability higher than 1/2.

Treatment	Low penalty				High penalty			
	Subjective probability			Total	Subjective probability			Total
	-1	0	1		-1	0	1	
BASE	9	25	6	40	13	26	1	40
PIC	44	40	20	104	23	21	10	54
Total	53	65	26	144	36	47	11	94
$\chi^2(2) = 7.0642$, $\text{Pr} = 0.029$					$\chi^2(2) = 8.7830$, $\text{Pr} = 0.012$			

Table 1: Distribution of subjective probability by treatment

In the BASE treatments more than 60% of the subjects (62.5% and 65% in the low and high penalty setup respectively) considered their probability for an audit to be equal to 1/2. But in the PIC treatments far more subjects answered that their subjective probability for being audited deviates from 1/2; the share of subjects reporting an audit probability of 50% is roughly 38% in both the low and the high penalty setup (see Table 1). Hence, for both the low and the high penalty setup the share of subjects reporting an subjective audit probability that is not equal to 50% increases by roughly 24-27 percentage points. When carrying out χ^2 -tests for the two subtables, the null hypothesis that the variables Treatment and Subjective audit probability are independent is rejected for both the low and high penalty setup ($p < 0.05$). Hence, the table supports the idea that the share of subjects in the PIC treatments who believe their probability deviates from 1/2 should be higher than in the BASE treatments.

We test Hypotheses 1A and 1B by estimating two logit models. The first

	(1)	(2)
	Lower than 1/2	Higher than 1/2
PIC	0.935** (0.311)	1.276** (0.462)
Constant	-0.841** (0.256)	-1.986*** (0.404)
N	201	149

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Share of subjects with subjective probability lower/higher than 1/2 (logit model)

column in Table 2 estimates the share of subjects reporting a subjective audit probability lower than 1/2. The predicted share is equal to 30.13% in the BASE treatments compared to 52.35% in the PIC treatments.⁷ This supports Hypothesis 1A with the estimate being significant at the 1% level. Column (2) predicts that the share of subjects reporting a subjective probability higher than 1/2 is equal to 12.06% in the BASE treatments and equal to 32.96% in the PIC treatments. This supports Hypothesis 1B with the estimate being significant at the 1% level.⁸

We now turn to Hypotheses 2A and 2B. Analyzing the descriptive statistics for the pooled sample (Table 3) supports the idea that the subjects form beliefs about their subjective audit probability and behave as predicted by the Hypotheses: if the penalty is low then compliance increases (left), and if the penalty is high then compliance decreases (right). As the descriptive

	Penalty	
	low	high
BASE	0.28	0.65
PIC	0.33	0.55

Table 3: Share of truthful (high-endowment) reports by treatment

evidence does not take into account that the reporting decisions from the same subject are correlated, we employ a logit model where the standard

⁷The predictions are calculated by inserting the coefficient of the constant for BASE and the coefficient of the constant plus the coefficient of PIC into the logistic cdf, respectively.

⁸In further (unreported) regressions, we entered a dummy indicating observations from the high penalty variants to control for a possible confounding by penalty. This check does not affect the main results and produces no evidence for such a confounding.

errors are clustered on the subject-level. The results obtained from the logit model (Table 4) suggest that Hypothesis 2B is supported by the data for the high-penalty variant ($p < 0.1$)⁹ but Hypothesis 2A is not supported by the data for the low-penalty variant ($p = 0.195$)¹⁰.

	(1)
PIC-low	0.245 (0.286)
BASE-high	1.580 (0.332)
PIC-high	-0.671 (0.431)
Constant	-0.943 (0.232)
N	1148

Standard errors clustered on the subject-level in parentheses

Low Penalty $H_0 : \text{PIC} \leq \text{BASE}$ vs. $H_1 : \text{PIC} > \text{BASE}$: $z = 0.86$, $p = 0.195$

High Penalty $H_0 : \text{PIC} \geq \text{BASE}$ vs. $H_1 : \text{PIC} < \text{BASE}$: $z = -1.31$, $p = 0.093$

The table presents results from a logit model where the response variable is a dummy variable that is equal to one if a high endowment is truthfully reported and zero otherwise. The explanatory variables are a series of dummy variables indicating the respective treatment in which the subject participated.

Table 4: Compliance / Reporting truthfully by treatment (logit model)

To summarize this analysis, we find evidence for differences in subjective audit probabilities for the BASE treatments compared to the PIC treatments. A considerable share of subjects forms subjective probabilities that differ from the average audit probability if they know that the audit choice is potentially influenced by their picture. In the PIC treatments the subjective probabilities deviate systematically from 1/2 for a large share of the subjects. Here we focused on qualitative deviations from 1/2 and the effects also measure potentially small deviations from 1/2. We also confirmed that the subjective audit probability in the PIC treatments may exceed the average audit probability for some subjects, and may fall short of the average audit probability for another share of subjects.

Quantitatively large deviations from 1/2 should also induce differences in compliance decisions between BASE and PIC. We used two compliance treatments to elicit and identify such behavioral changes. The effects we measure point into the direction of Hypotheses 2A and 2B. However, the

⁹ $H_0 : \text{PIC-high} \geq \text{BASE-high}$ vs. $H_1 : \text{PIC-high} < \text{BASE-high}$: $z = -1.31$, $p = 0.093$

¹⁰ $H_0 : \text{PIC-low} \leq \text{BASE-low}$ vs. $H_1 : \text{PIC-low} > \text{BASE-low}$: $z = 0.86$, $p = 0.195$

effects are statistically weak. There may be a number of reasons why the effect on subjective probabilities does not show up in behavioral changes more strongly. One reason is the order of magnitude that was necessary for inducing a behavioral change. To tilt the monetary incentive of whether to report truthfully or to underreport the deviation of subjective probability from the average probability of $1/2$ had to be substantial: 16.7 percentage points or more.¹¹ Nevertheless, we find some mild evidence suggesting that the compliance behavior differs between the BASE treatment and the PIC treatments, and the direction of the effects is in line with Hypotheses 2A and 2B, with the treatment effect in the high penalty variant being statistically significant at a conventional level ($p < 0.1$).¹²

4 Conclusion

If tax officers base their audit decisions on a comparison of pictures of the set of subjects who potentially underreported, this matters for the subjects who make compliance decisions. The fact that a photo is used to assess their truthfulness affects the subjective probability they form about whether they will receive an audit. Some individuals think that this increases the probability for an audit. A larger share of individuals thinks that this decreases the probability for an audit since they believe that their picture looks more trustworthy than average. This change in subjective audit probabilities may affect the compliance decisions, as it affects the expected cost/benefit from underreporting. This effect shows up in the data. If the penalty for underreporting is low then compliance increases slightly if a picture is shown. However, if the penalty is high, compliance decreases significantly. Overall, the experiment reveals that subjects think that the ability to detect their lies is not necessarily tied to their actual performance (voice, body language etc.), but that possible auditors may base their decisions on some more permanent cues.

¹¹Sample size may also be an issue. As the expected effect size is unknown a priori, it is unclear which sample size is needed to carry out formal hypothesis tests. For the discussion on the (ab)use of applying uniformly a 5% level and related problems, see, e.g. McCloskey and Ziliak (1996) and Krämer (2011).

¹²This weak effect is in line with an analysis of the observed correlation between the subjective probability assessment and individuals' reporting decisions, which is small and insignificant.

References

- [1] Andreoni, J. and Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments, *Journal of Economic Psychology* 29(1): 73–93.
- [2] Castillo, M. and Petrie, R. (2010). Discrimination in the lab: Does information trump appearance?, *Games and Economic Behavior* 68(1): 50–59.
- [3] Coricelli, G., Joffily, M., Montmarquette, C. and Villeval, M. (2010). Cheating, emotions, and rationality: an experiment on tax evasion, *Experimental Economics* 13(2): 226–247.
- [4] DeBruine, L. M. (2002). Facial resemblance enhances trust, *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269(1498): 1307–1312.
- [5] Eckel, C. C. and Petrie, R. (2011). Face value, *American Economic Review* 101(4): 1497–1513.
- [6] Greiner, B. (2004). An online recruitment system for economic experiments, MPRA Paper 13513, University Library of Munich, Germany.
- [7] Habyarimana, J., Humphreys, M., Posner, D. and Weinstein, J. (2007). Why does ethnic diversity undermine public goods provision?, *American Political Science Review* 101(4): 709–725.
- [8] Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects, *American Economic Review* 92(5): 1644–1655.
- [9] Konrad, K. A., Lohse, T. and Qari, S. (2012). Customs compliance and the power of imagination, CESifo Working Paper Series 3702, CESifo Group Munich.
- [10] Krämer, W. (2011). The cult of statistical significance—what economists should and should not do to make their data talk, *Schmollers Jahrbuch* 131: 455–468.
- [11] McCloskey, D. and Ziliak, S. (1996). The standard error of regressions, *Journal of Economic Literature* 34(1): 97–114.
- [12] Mobius, M. M. and Rosenblat, T. S. (2006). Why beauty matters, *American Economic Review* 96(1): 222–235.

- [13] Solnick, S. J. and Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions, *Organizational Behavior and Human Decision Processes* 79(3): 199–215.
- [14] Wilson, R. K. and Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game, *Political Research Quarterly* 59(2): 189–202.