Experimental Studies on Tax Compliance Behavior

Inauguraldissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität zu Köln

2019

vorgelegt von

Nadja Marion Müller, M.Sc.

aus Mechernich

Referent: Prof. Dr. Martin Fochmann, Universität zu Köln

Korreferent: Prof. Dr. Michael Stich, Universität zu Köln

Tag der Promotion: 16. Dezember 2019

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Lehrstuhl für Behavioral Accounting/Taxation/Finance der Universität zu Köln. Sie wurde im Juni 2019 von der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln als Dissertation angenommen. Das Entstehen dieser Arbeit wurde durch die Unterstützung zahlreicher Personen geprägt, denen ich an dieser Stelle ganz herzlich danken möchte.

Mein tiefster Dank gilt meinem Doktorvater Herrn Prof. Dr. Martin Fochmann für seine hilfsbereite Unterstützung und seine uneingeschränkte Bereitschaft, sein Wissen mit mir zu teilen. Seine wertvollen Ratschläge, geduldige Nachsicht und wohlwollende Förderung haben die Promotion für mich erst möglich gemacht. Ich danke ihm von Herzen für das mir entgegengebrachte Vertrauen und seine Begleitung als Lehrer und als Freund.

Zudem danke ich Herrn Prof. Dr. Michael Stich für die Erstellung des Zweitgutachtens und Herrn Prof. Dr. Christoph Kuhner für die Übernahme des Vorsitzes der Prüfungskommission. Für die tolle Zusammenarbeit möchte ich auch Nadja Fochmann und Herrn Prof. Dr. Michael Overesch danken, in deren Koautorenschaften Beiträge dieser Dissertation entstanden sind.

Bedanken möchte ich mich auch bei Manuel Schwaninger, der mich mit seiner Neugier und seiner Freude an der Forschung angesteckt und stets unterstützt hat. Durch ihn habe ich während meiner Promotionszeit nie die Motivation verloren. Danken möchte ich zudem meinen Kolleginnen und Kollegen an der Universität zu Köln für die freundschaftliche Arbeitsatmosphäre und viele gemeinsame Aktivitäten außerhalb der Universität. Dank gebührt auch Christine Straberg, die mich als studentische Hilfskraft bei der Durchführung der Experimente verlässlich unterstützt hat. Besonders bedanken möchte ich mich bei Gary Rüsch, André Hoppe und Selina Orthaus, auf deren Unterstützung ich mich jederzeit verlassen konnte und deren Freundschaft ein wertvoller Gewinn für mich ist.

Darüber hinaus gebührt ganz besonderer Dank meiner Familie. Durch ihre stete und bedingungslose Unterstützung haben meine Mutter Annemie Müller und meine Großeltern Katharina und Helmut Meyer wesentlich zum Gelingen dieser Arbeit beigetragen. Ihr liebevoller Rückhalt und unerschöpflicher Zuspruch ermöglichen es mir, jeden Weg in meinem Leben gehen zu können.

Köln, im Dezember 2019 Nadja Müller

Content

Chapter 1

Introduction

Chapter 1

1.1 Motivation and Object of Research

Ever since, taxation is a topic of central relevance for states. Usually, tax revenues build the financial foundation for governments, and individuals and organizations are obliged to pay taxes by a state's tax laws. For policy makers, a pivotal question is: *How to ensure that taxpayers pay the correct amount of taxes, i.e., act tax compliant?* Extensive work is done by governments to elaborate policy proposals covering various aspects of the tax declaration process in order to increase tax compliance. Also globally, serious effort is undertaken to establish joint work of governments and develop global solutions to this question. In the last years, for example, the OECD started several comprehensive projects to coordinate work on improving tax compliance.[1]

But, how effective are these measures? In order to find an answer to the question "*How to increase tax compliance?"* it is crucial to understand the effects of different policy measures on tax compliance. And in order to understand these effects, it is essential to better understand tax compliance behavior.

The sole reliance on standard economic theory to predict compliance behavior has not led to a satisfying answer. Researchers observe that standard economic theory is not able to predict human decision making accurately. One central assumption in the theory is that decision makers act fully rational. But what we observe in reality is: they do not.[2] To improve the predictions of decision-making, behavioral economists relax some of the theoretical assumptions and incorporate psychological insights into economic theory (e.g., bounded rationality, limited self-interest, social preferences).[3] Psychological insights are of essential importance as economic factors alone do not explain behavior. Behavioral economics aims to close the gap between economics and psychology and to provide further important insights into the determinants of decision-making.

Experiments offer a valuable instrument to identify and study additional factors that might influence tax compliance. As Magnus Johannesson stated in his speech introducing the winner of the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2017, evidence from lab and field experiments is needed to complement theory.[4] The evidence from

---

1 For example, the "International Compliance Assurance Programme" (OECD, 2018), the project "The Changing Tax Compliance Environment and the Role of Audit" (OECD, 2017) and the project "Co-operative Tax Compliance - Building Better Tax Control Frameworks" (OECD, 2016).

2 See, for example, Kahneman, and Tversky (1979), Thaler (1980), Kahneman et al. (1982), Simon (1982), Laibson (1997), Camerer (1998), Ariely (2008).

3 See Camerer and Loewenstein (2003) and Thaler (2016) for comprehensive overviews.

4 Introduction to the Prize Lecture of Richard H. Thaler at Stockholm University, December 2017. Richard H. Thaler received the prize for his contributions to Behavioral Economics.

experiments can be used to refine and improve the theoretical models. In order to derive useful policy implications, we need to know *how* decision-makers behave, and *why* they behave in that way. The central question in behavioral economics is: *What drives human behavior?* Experimental research aims to contribute to the understanding of decision-making and to expand the predictive power of economic theory. Policy measures are not effective if based on economic theory that fails to predict decision-makers' behavior. In order to design expedient policy measures, the insights provided by experimental research are of crucial importance.

In recent years, an increasing number of governments have recognized the importance of insights provided by experimental research for the design of effective and targeted policy measures and established separate behavioral governmental units.[5] These units incorporate psychological insights into the design of policy measures, conduct experiments to study the effects of proposed policy measures and evaluate their usability.

With their research on the *nudge* concept, Richard H. Thaler and Cass R. Sunstein (2008) highlight the relevance of psychological phenomena for the prediction of decision making. They incorporate psychological assumptions to predict human behavior and show that behavior can be changed by implementing nudges, i.e., small changes to the environment without changing the economic incentives in that situation. Based on their insights, they offer policy recommendation in various areas, such as health, environment, finance, education.

Experimental studies can offer surprising results. There are several examples, where researchers observe a different behavior as expected beforehand. For example, as Denmark introduced the prefilling of tax-deductible charitable contributions in taxpayers' tax returns, one aim was to limit the possibilities of untruthful reporting, which should lead to lower deductions claimed and higher tax revenues. Contrary to their expectations, Gillitzer and Skov (2016) observe that the number of tax deductions claimed doubles, resulting in a substantial decrease of tax revenues in the first year of prefilled charitable contributions in Denmark. This example highlights that policy measures do not necessarily lead to the theoretically expected result and that in order to derive effective policy measures, it is essential to take into account insights provided by experiments.

This thesis focuses on the behavioral perspective of tax compliance. The standard economic model predicting tax compliance is solely based on monetary factors (Becker, 1968,

[5] For example, the federal government of Germany implemented the unit "wirksam regieren" (governing effectively) and the U.S. government established the "Social and Behavioral Sciences Team". Also the European Commission implemented an in-house "Joint Research Centre" to deal with behavioral insights relevant for policy.

Allingham and Sandmo, 1972, Yitzhaki, 1974). The main monetary factors determining compliance are the audit probability, the penalty in case of detected non-compliance, and the tax rate. The literature has provided extensive research on how these factors can influence tax compliance behavior (see Alm et al., 1995, Andreoni et al., 1998, Torgler, 2002, Hofmann et al., 2008, Alm, 2012, and Slemrod, 2016 for comprehensive literature reviews). But this standard model predicts a significantly lower tax compliance level than is observed in the real world. Thus, there need to be other factors influencing tax compliance behavior. In recent years, research has incorporated non-pecuniary factors, such as social norms, fairness and moral costs of dishonest behavior (e.g., Kim, 2003, Fortin et al., 2007, Dulleck et al., 2016).

With the following three essays on compliance behavior, I aim to contribute to the literature that studies the non-monetary factors influencing decision-making. The first two essays examine one very current topic in taxation: the prefilling of tax returns. As a consequence of the digital transformation, vast amounts of data can be collected, processed and analyzed. The digital data exchange allows tax authorities to receive data on the taxpayer from, for example, employers, banks, and social insurance agencies, and this information in turn can be prefilled for the taxpayer in her tax forms. While Scandinavian countries such as Finland, Sweden or Denmark, already implemented prefilled tax returns in the 1990s, an increasing number of countries have implemented prefilled tax returns lately, for example, Germany, Canada, Australia, Spain and the UK (OECD, 2017a). With prefilling tax returns for the taxpayer, states aim at supporting the taxpayer in filing their tax returns and provide an image of service orientation and cooperation. At the same time, the data exchange allows tax authorities to have more control over the tax relevant information of a taxpayer, and aims at improving tax compliance (OECD, 2017a).

Surprisingly, so far there is very little evidence on the effects of prefilling. Some studies provide first important insights using empirical field data of tax authorities on third-party reporting. Kleven et al. (2011) provide initial empirical results that prefilling can have compliance consequences and find that tax compliance is higher for income that is subject to third-party reporting (already prefilled) than for self-reported income (no prefilling). In contrast, Gillitzer and Skov (2016) and Kotakorpi and Laamanen (2016) observe that partly prefilled tax returns have either no effect or even a negative effect on reported taxable income. Although an increasing number of countries has introduced prefilling in the tax declaration process, aiming to increase tax compliance, “the direct effect of pre-population on compliance is unclear” (Slemrod, 2016, p. 64).

I aim to fill this research gap with two experimental studies on the effects of prefilled tax returns on compliance behavior. I conduct lab experiments in order to identify and analyze the effects prefilling might have on compliance behavior in a controlled environment. The first study "Less Cheating? The Effects of Prefilled Forms on Compliance Behavior" is joint work with Prof. Dr. Martin Fochmann, Chair of Behavioral Accounting/Taxation/Finance at the University of Cologne, and Prof. Dr. Michael Overesch, Chair of Business Taxation at the University of Cologne. In a controlled experiment, we investigate how fully prefilled tax forms influence tax compliance behavior. Our joint work was characterized by strong cooperation and a constant exchange of ideas. Among others, I contributed to designing and programming the experiment, developing the hypotheses, providing a literature overview, conducting the experiment, implementing empirical analyses and writing the paper. The paper was presented, for example, in the PhD course "Advanced Experimental Economics" of Matthias Sutter in Cologne in 2016, the Doctoral Research Seminar in Taxation in Cologne in 2017, the arqus (work group for quantitative taxation) Conference 2017 in Berlin, the GfeW (German Association for Experimental Economic Research) Conference 2017 in Kassel, the European Accounting Association (EAA) Annual Congress 2018 in Milan and the Annual Congress of the International Institute of Public Finance (IIPF) 2018 in Tampere .

Building on the insights of this first study, the second study "Partial Prefilling: Compliance Behavior und Adjustment Behavior for Prefilled Values" is a single author paper and has been created entirely under my own responsibility. The study aims to test the findings of the first study in different prefilling scenarios. If tax returns are prefilled, it might often be the case that they are not completely prefilled. For example, in the case of third party reporting, information from employers, banks and social insurance agencies are available and can be prefilled, but other types of income sources, such as income from renting property or from self-employed activities, need to be reported by the taxpayer herself. It is important to study the effects of prefilling on tax compliance behavior in the different prefilling situations that taxpayers might face. Thus, my first and second study contribute to the literature on tax compliance behavior and the insights gained in the experiments might be of practical relevance for tax administrations and offer policy implications for governments.

The third essay focuses on another important aspect of tax compliance behavior. Tax compliance is not only of interest in individual tax filing situations, but also when considering institutions and organizations. Within larger organizations, decisions are rarely taken by one single individual, but rather by, for example, teams, departments, boards, or committees. But can the compliance decision of a group be simply seen as an aggregation of individual

behaviors? There might be additional factors that influence group compliance behavior. So far, the literature on tax compliance has focused on individual tax compliance behavior. Looking further, one finds that the literature studying differences in decision making of groups and individuals focuses either on immoral behavior (Conrads et al., 2013, Kocher et al., 2017, Sutter, 2009) or on risk-taking behavior (Fahr and Irlenbusch, 2011, Feri et al., 2010, Sutter, 2007). There is no overlap of these two parts of the literature. The third study "Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups" aims to fill this research gap and investigates the differences in tax compliance behavior between groups and individuals. This study is joint work with Prof. Dr. Martin Fochmann, Chair of Behavioral Accounting/Taxation/Finance at the University of Cologne, and Nadja Wolf, Research Assistant at the Chair of Business Taxation at the University of Hannover. Our joint work was characterized by great collaboration and complementary work. Among others, I contributed to designing the experiment, developing the hypotheses, providing a literature overview, conducting the experiment, implementing empirical analyses and writing the paper. The paper was presented, for example, at the Thurgau Experimental Economics Meeting (theem) 2018 in Konstanz.

1.2 Less Cheating? The Effects of Prefilled Forms on Compliance Behavior

The first essay "Less Cheating? The Effects of Prefilled Forms on Compliance Behavior" (joint work with Prof. Dr. Martin Fochmann and Prof. Dr. Michael Overesch) investigates how fully prefilled forms influence tax compliance behavior. We conduct a laboratory experiment with 213 participants to study the effects of prefilling on tax compliance behavior. In the experiment, we keep all monetary factors that influence compliance behavior, such as the tax rate, audit probability and fines, constant to exclude that compliance behavior might differ due to monetary factors. We argue that prefilling might influence tax compliance via several psychological phenomena in different prefilling situations. Our experimental design allows us to provide a clean test of how prefilling affects tax compliance behavior and provides a high level of internal validity. Furthermore, we are able to clearly distinguish between correctly and incorrectly prefilled tax returns and, thus, can isolate their respective influences on tax compliance. Our set-up additionally allows us to study how individuals adjust prefilled tax returns and how adjustment behavior varies between different prefilling scenarios.

We show that correct prefilling enhances compliance. Subjects are significantly more tax compliant when the tax return is correctly prefilled compared to when the tax return is not

prefilled (and thus, needs to be filled in by the taxpayer herself). However, in cases of incorrect prefilling, we observe asymmetric effects. If prefilled income is lower than true income, we find no positive effect on compliance, and compliance is on the same level as with blank forms. If prefilled income is higher than true income, prefilling still has a positive effect on compliance. In that case, compliance is on the same level as with correctly prefilled forms and higher than with blank forms. Thus, in the case that taxpayers would benefit from incorrect prefilling, prefilling has no effect on the tax compliance level when compared with blank tax returns. In contrast, if taxpayers would suffer from incorrect prefilling, prefilling still has a positive effect on compliance compared to the case of blank tax returns.

Furthermore, we find that individuals are aware of the incorrectly prefilled values and actively adjust them in their tax returns. More than ninety percent of the incorrect values are adjusted by participants. The adjustment behavior depends on the type of prefilling error. Prefilled income higher than true income is almost always adjusted downward. For prefilled income below true income, however, adjustments depend on the deviation from true income. If the deviation is small, subjects do not adjust prefilled values, but if the deviation is large, subjects adjust lower prefilled values upward.

Our study contributes to the literature by studying the effects of prefilling on compliance behavior, which have not yet been studied in a laboratory experiment. Individuals and decision makers in organizations and institutions are often confronted with prefilled data and prefilled forms in almost all industries. Our results suggest that in situations where dishonesty is of concern, the presence of defaults and the quality of these defaults both influence dishonest behavior.

Our results also provide an answer to the following question: Should we make use of prefilled forms or prefilled entry masks? Indeed, prefilled forms reduce compliance costs (Goolsbee, 2006, Klun, 2009, Evans and Tran-Nam, 2010). However, what about compliance behavior on an aggregated level? Our results suggest that prefilling does not always outperform non-prefilling in every situation but that prefilling is preferable over non-prefilling on an aggregated level.

Furthermore, as prefilling tax returns is a service that the tax administration provides to taxpayers, our study is related to the service paradigm literature. For example, Alm et al. (2010) show that services from the tax administration (e.g., agency-provided information) have a positive and significant impact on compliance behavior. Our paper also contributes to the slippery slope literature (Kirchler et al., 2008) and to the trust paradigm literature (Alm and Torgler, 2011). Both bodies of literature show that tax compliance is influenced not only by the

power of the tax authority to conduct audits, collect taxes and punish tax evasion (enforcement paradigm) but also by the trust in authority. While an increase in the power of authority leads to higher enforced tax compliance, an increase in the trust in authority leads to higher voluntary tax compliance (Wahl et al., 2010, Kastlunger et al., 2013, Kogler et al., 2013). If tax returns are correctly prefilled by the tax authority, the trust in authority might increase, and therefore tax compliance might increase as well. However, if tax returns are incorrectly prefilled, the trust in authority might be lower, leading to a decrease in tax compliance.

1.3 Partial Prefilling: Compliance Behavior und Adjustment Behavior for Prefilled Values

The second essay "Partial Prefilling: Compliance Behavior und Adjustment Behavior for Prefilled Values" investigates the influence of partly prefilled tax returns on compliance behavior. Building on the first essay „Less Cheating? The Effects of Prefilled Forms on Compliance Behavior" which shows that fully prefilled forms have a significant influence on compliance behavior, the second essay aims to replicate these findings in different prefilling scenarios, and to provide additional analyses on the compliance behavior and the adjustment behavior for prefilled values.

To test how partly prefilled tax returns influence tax compliance, I conduct a lab experiment with 158 participants. In the experiment, individuals receive a tax return, in which half of the information is already prefilled, and the other half is not prefilled, thus needs to be filled in by the taxpayer herself. I consider the following cases: a tax return can be partly prefilled with correct information or partly prefilled with incorrect information, whereas the incorrect information can either lead to tax savings or tax disadvantages for the taxpayer. Additionally, I investigate the case that one tax return contains different kinds of errors, i.e., errors that lead to lead to tax savings and errors that lead to tax disadvantages. As in the first study, all monetary aspects are kept constant, and the only variation is the prefilling in the treatments.

I find that partial prefilling with correct values increases compliance for the prefilled fields compared to the non-prefilled fields. Consequently, correct prefilling enhances compliance not only in the case of fully prefilled tax returns, but also in the situation that forms are only partly prefilled. In case of incorrect prefilling, I do not observe a difference in compliance for the prefilled fields compared to the non-prefilled fields, independent of whether the error results in tax savings or tax disadvantages for the taxpayer.

Additionally, I show that the findings of the first study regarding the adjustment behavior for prefilled values can be transferred to different prefilling scenarios. The results show that the amount of adjustment of incorrectly prefilled values purely depends on the direction of the prefilling error and its size, i.e., the deviation of the incorrect value from the true value. Thus, the data suggests that independent of the prefilling scenario, individuals reveal the same adjustment behavior for incorrectly prefilled values.

My study contributes to the literature by providing insights on the effects of partial prefilling on individual compliance behavior, which have not yet been studied in a laboratory experiment. Tax administrations might not always have access to all tax-relevant information on a taxpayer, and in that case only partial prefilling is feasible. Therefore, it is important to study how the original findings regarding the effects of fully prefilled forms on compliance behavior can be transferred to other prefilling scenarios. The results might also have practical implications for governments. The results suggest that correct prefilling can enhance compliance also in the case that only partial information is available, but in situations where errors in prefilling occur, tax administration should not expect differences in compliance compared to non-prefilling. Nevertheless, the results suggest that prefilling is preferable over non-prefilling.

Second, my study provides valuable insights on the adjustment behavior for prefilled values. Although the digital processes of data exchange should be accurate and trustworthy, errors in these processes might occur, and these errors could lead to tax savings or tax disadvantages for the taxpayer, if not adjusted. I contribute to the understanding of tax compliance behavior by providing further insights on the adjustment behavior of prefilled - especially incorrectly prefilled - values in different prefilling scenarios. The findings might be of practical relevance for tax administrations, especially in situations where errors are likely to occur, as my findings suggest that independent of the overall prefilling situation of the taxpayer, taxpayers reveal the same adjustment behavior for incorrectly lower (higher) values.

Additionally, my study contributes to the literature on replication in research. In recent years, researchers in all areas of science emphasize the need for replication studies and highlight their importance for scientific reliability (see, for example, Rosenblat et al., 2015, Camerer et al., 2016). I contribute to the literature by providing a replication study which examines how the findings of the first study can be transferred to different prefilling scenarios.

1.4 Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups

The third essay "Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups" (joint work with Prof. Dr. Martin Fochmann and Nadja Wolf) studies the differences in compliance decisions between individuals and groups. We conduct a controlled experiment with 189 participants and base our design on the cheating framework of Kocher et al. (2017). This cheating framework is extended by a risk dimension: unethical behavior can be disclosed and penalized. Thus, we use a tax compliance context to study the risk and the honesty dimension simultaneously. Additionally, we identify and discuss some of the potential reasons for collective non-compliant behavior.

In the experiment, participants face the situation of being an employee of a tax department that is responsible for filing the annual income tax return of the organization. Declaring less than the true income saves taxes at the organizational level. When non-compliance is disclosed by an audit, the organization has to repay the evaded taxes plus a penalty. In the individual setting, the tax department consists of one employee deciding on the tax compliance level of the organization. In the group setting the tax department consists of three employees. The three individuals in the group setting first communicate in an anonymous group chat before voting determines the group decision. Employees receive a fixed remuneration and a variable remuneration that depends on the company's after tax profit. Thus, a tax department employee gains from tax evasion if no audit occurs and experiences a disadvantage if an audit occurs. In our setup, we ensure that the monetary payoff for one employee is identical in the individual and group setting.

We observe that compliance is significantly lower in the group than in the individual setting. Consequently, groups are less compliant than single decision makers. Furthermore, we are able to show that risk plays a substantial role when individuals make compliance decisions in groups. Risk arguments are the arguments most frequently mentioned in group communication and, more importantly, the influence of communication on the group's compliance behavior is mainly driven by arguments relating to risk. This provides strong evidence that risk concerns are non-negligible and have a high impact on dishonest behavior. Our results suggest that communication in groups alters the risk tolerance of the group members. The frequent exchange of risk arguments encouraging non-compliance changes the norm perception of individuals and increases the willingness to enter in risky non-compliance behavior. This is in line with the literature suggesting than learning about preferences of others

shifts norm perception (Gino et al., 2009, Kocher et al., 2017) and provides evidence for a conformism effect (Kocher et al., 2013, Lahno and Serra-Garcia, 2015).

Additionally, we find that group interaction induces a negative spill-over effect on subsequent individual compliance. The shift in norm perception through group interaction is a sustaining effect and observable in a significantly lower individual compliance level after group interaction compared to individual compliance before group membership. However, the compliance level in the individual setting after group interaction is significantly higher than in the preceding group setting. This finding suggests that the shift in norm perception is not the only driver for the difference in behavior between groups and individuals in our setup. Otherwise we would have observed the same compliance level in the group setting and in the subsequent individual setting. As the increase in compliance is not observed in a setting exclusively focusing on the honesty dimension (Kocher et al., 2017), our finding suggests that the effect of group membership on behavior is stronger than the honesty literature without risk concerns predicts. This emphasizes the relevance of the risk dimension in group decision making.

Lastly, a categorization of subjects indicates that individuals react differently to group interaction and thus, are heterogeneous in their behavior. However, differences are rather systematic and nearly all individuals can be assigned to three types of decision makers.

We contribute to the literature by studying the differences in compliance decisions between individuals and groups. The compliance literature so far has focused on studying individual compliance decisions. But decisions in organizations are often taken in groups rather than by one single individual. Although organizations and their managers should act in a way that is compliant with the law, there are several examples of scandals where non-compliant behavior was detected. For example, Lux Leaks and the Panama Papers made public a large number of organizations involved in tax fraud.[6] Also corruption (e.g., Walmart, Siemens, Halliburton, KBR), financial accounting fraud (e.g., Enron, Worldcom, Tyco), insurance fraud (e.g., Hospital Corporation of America, TAP Pharmaceuticals) and the Volkswagen's diesel emissions manipulation are examples for detected non-compliant behavior of organizations. If this behavior is disclosed, negative consequences such as penalty payments, compensation

[6] Lux Leaks is the release of secret tax documents in 2014 on tax deals of Luxembourg's tax authority with multinational corporations by the International Consortium of Investigative Journalists (ICIJ) (for details see IJIC (2014): https://www.icij.org/investigations/luxembourg-leaks/). The Panama Papers are a leak of secret documents on the exploitation of offshore tax regimes from an anonymous source by the German newspaper "Süddeutsche Zeitung" which analyzed them in cooperation with the IJIC (for details see Süddeutsche Zeitung (2016): https://panamapapers.sueddeutsche.de/articles/56febff0a1bb8d3c3495adf4/).

claims, loss of reputation, or court hearings can occur for the organization and for the involved managers.

There is a large literature comparing decision making of groups and individuals with the focus either on immoral behavior (Conrads et al., 2013, Kocher et al., 2017, Sutter, 2009) or on risk-taking behavior (Fahr and Irlenbusch, 2011, Feri et al., 2010, Sutter, 2007). However, there is no overlap of both parts of the literature: dishonest behavior is studied without the risk of negative monetary consequences and risk-taking behavior is studied in situations without ethical concerns. In our tax compliance setup, non-compliant behavior is both immorally and risky. Thus, we contribute to the literature by connecting both parts of the literature.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323–338.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1), 54–77.

Alm, J., Cherry, T., Jones, M. and McKee, M. (2010). Taxpayer information assistance services and tax compliance behavior. *Journal of Economic Psychology*, 31(4), 577–586.

Alm, J., Sanchez, I. and De Juan, A. (1995). Economic and noneconomic factors in tax compliance. *Kyklos*, 48, 3–18.

Alm, J. and Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635–651.

Andreoni, J., Erard, B. and Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, 36, 818–860.

Ariely, D. (2008). Predictably irrational. Harper Collins.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 169–217.

Camerer, C. F. (1998). Bounded rationality in individual decision making. *Experimental economics*, 1(2), 163–183.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.

Camerer, C. F., and Loewenstein, G. (2003). Behavioral economics: Past, present, future.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology* 34, 1–7.

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M. and Torgler, B. (2016). Tax compliance and psychic costs: behavioral experimental evidence using a physiological marker. *Journal of Public Economics*, 134, 9–18.

Evans, C. and Tran-Nam, B. (2010). Managing tax system complexity: building bridges through prefilled tax returns. *Australian Tax Forum*, 25, 245−274.

Fahr, R., and Irlenbusch, B. (2011). Who follows the crowd – Groups or individuals? *Journal of Economic Behaviour & Organization,* 80, 200–209.

Feri, F., Irlenbusch, B., and Sutter, M. (2010). Efficiency Gains from Team-based Coordination – Large-Scale Experimental Evidence. *American Economic Review,* 100, 1892–1912.

Fortin, B., Lacroix, G. and Villeval, M. C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11-12), 2089–2112.

Gillitzer, C. and Skov, P. E. (2016). The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark. EPRU Working Paper Series.

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior. *Psychological Science*, 20, 393–398.

Goolsbee, A. (2006). The simple return: Reducing America's tax burden through return-free filing. Brookings Institution Hamilton Project Discussion Paper 2006-04.

Hofmann, E., Hoelzl, E. and Kirchler, E. (2008). Preconditions of voluntary tax compliance: Knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. *Journal of Psychology*, 216, 209–217.

Kahneman, D., Slovic, S. P., Slovic, P., and Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. Cambridge university press.

Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.

Kastlunger, B., Lozza, E., Kirchler, E. and Schabmann, A. (2013). Powerful authorities and trusting citizens: The Slippery Slope Framework and tax compliance in Italy. *Journal of Economic Psychology*, 34, 36–45.

Kim, Y. (2003). Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. *Journal of Public Economics*, 87(7-8), 1591–1616.

Kirchler, E., Hoelzl, E. and Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, 29, 210–225.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.

Klun, M. (2009). Prefilled income tax returns: Reducing compliance costs for personal income taxpayers in Slovenia. *Financial Theory and Practice*, 33(2), 219–233.

Kocher, M. G., Pogrebna G., and Sutter, M. (2013). Other-regarding preferences and management styles. *Journal of Economic Behaviour & Organization*, 88, 109–132.

Kocher, M. G., Schudy, S., and Spantig, L. (2017). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science*, 64, 3971–4470.

Kogler, C., Batrancea, L., Nichita, A., Pantya, J., Belianin, A. and Kirchler, E. (2013). Trust and power as determinants of tax compliance: Testing the assumptions of the slippery slope framework in Austria, Romania and Russia. *Journal of Economic Psychology*, 34, 169–180.

Kotakorpi, K. and Laamanen, J. P. (2016). Prefilled income tax returns and tax compliance: Evidence from a natural experiment. Tampere Economic Working Paper 104.

Lahno, A. M., and Serra-Garcia, M. (2015). Peer effects in risk taking: Envy or conformity? *Journal of Risk and Uncertainty*, 50, 73–95.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443–478.

OECD (2018). International Compliance Assurance Programme Pilot Handbook. OECD Publishing.

OECD (2017). The Changing Tax Compliance Environment and the Role of Audit. OECD Publishing.

OECD (2017a). Third-party data management – the journey from post-assessment crosschecking to pre-filling and no-return approaches. In: Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies. OECD Publishing.

OECD (2016). Co-operative Tax Compliance: Building Better Tax Control Frameworks. OECD Publishing.

Rosenblat, T., Fatas, E. and Deck, C. A. (2015). Replication in Experimental Economics. Emerald Group Publishing.

Simon, H. A. (1982). Models of bounded rationality. Cambridge MIT Press.

Slemrod, J. (2016). Tax compliance and enforcement: New research and its policy implications. Ross School of Business Working Paper No. 1302.

Sutter, M. (2007). Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. *Economic Letters,* 97, 128–132.

Sutter, M. (2009). Deception through telling the truth? Experimental evidence from individuals and teams. *Economic Journal,* 119, 47–60.

Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review*, 106(7), 1577–1600.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39–60.

Thaler, R. H., and Sunstein, C. R. (2008). Nudge: improving decisions about Health, Wealth, and Happiness. Yale University Press.

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

Wahl, I., Kastlunger, B. and Kirchler, E. (2010). Trust in authorities and power to enforce tax compliance. An empirical analysis of the "Slippery Slope Framework". *Law & Policy*, 32, 383–406.

Yitzhaki, S. (1974). Income tax evasion: A theoretical analysis. *Journal of Public Economics,* 3(2), 201–202.

Chapter 2

Less Cheating? The Effects of Prefilled Forms on Compliance Behavior

Less Cheating? The Effects of Prefilled Forms on Compliance Behavior

Martin Fochmann, Nadja Müller, Michael Overesch

University of Cologne

Abstract

As a consequence of digital transformation, individuals are often confronted with prefilled forms or prefilled data entry masks. In situations where cheating and lying are of concern, prefilling and defaults might reduce dishonest behavior. In a controlled experiment, we investigate how correctly and incorrectly prefilled forms influence tax compliance behavior. We show that correct prefilling enhances compliance. However, in cases of incorrect prefilling, we observe asymmetric effects. If prefilled income is lower than true income, we find no positive compliance effect, and compliance is on the same level as with blank forms. If prefilled income is higher than true income, prefilling still has a positive effect on compliance. In that case, compliance is on the same level as with correctly prefilled forms and higher than with blank forms. Our study contributes to the literature on cheating and lying by showing that prefilled forms and defaults affect compliance.

Keywords

Dishonesty, Defaults, Prefilled Forms, Tax Compliance, Behavioral Economics

JEL-Classification

C91, D14, H26

Acknowledgements: We thank Matthias Sutter, Kay Blaufus, Abdolkarim Sadrieh, Christoph Engel, Lisa Bruttel, Pete Lunn, Frank Hechtner, Peter Mohr, Jochen Hundsdoerfer, Rainer Niemann, Manuel Schwaninger, Florian Stolley, Christine Straberg and participants of the 2019 VHB Conference (Kommissionstagung), 2018 ZEW Conference, 2018 IIPF Conference, 2018 ESA Conference, 2018 EAA Conference, 2018 VHB Conference, 2018 Nudging and Information Conference, 2017 GfeW Conference, 2017 arqus Conference, and research seminars at the University of Potsdam and Max Planck Institute for Research on Collective Goods in Bonn for helpful comments and suggestions.

Chapter 2

Tables

Figures

2.1 Introduction

As a consequence of the digital transformation, massive amounts of data are collected, processed and analyzed in almost all industries and life situations. One benefit of this process is that individuals (e.g., taxpayers, managers, employees) automatically receive relevant information from an information system that supports their decision making. This unprecedented availability of and access to information has not only changed workflows and decision making in organizations and institutions, it has also changed how individuals complete forms and data entry masks. One important example is the filing of tax returns. If cheating and lying are of concern in such situations, prefilling and defaults might reduce dishonest behavior. However, the effect of prefilled forms or prefilled data entry masks on compliance behavior is largely unexplored, even though prefilling has become more and more common in various situations, such as the prefilling of tax returns, the claiming of travel expenses by employees, the reporting of insured events by policy holders or the collection of financial information from subdivision managers. In a controlled experiment, we investigate how prefilled forms affect compliance behavior, focusing specifically on how incorrectly prefilled values influence compliance.

In our investigation, we consider the situation of filing the annual income tax return. Today, taxpayers often start their tax declarations with prefilled forms. Customers who use tax software or file their tax returns online (with e-filing services) find that electronic tax return programs usually carry over the previous year's values to the subsequent year as an orientation aid. Consequently, a current tax return is initially prefilled with last year's numbers (e.g., salary, business income, expenses, deductions, tax credits). However, prefilled data is often incorrect, as data carried forward from the previous year does not fully reflect the conditions of the current year. Furthermore, prefilling plays a significant role in the work of revenue bodies in many countries, including the UK, Canada, Australia, Germany, France, Italy, Spain and Sweden (OECD, 2017). Automatic data exchanges between the tax authority and employers, social insurance agencies and banks enable systems to create tax returns that are prefilled before they are sent to taxpayers. Consequently, instead of completing blank forms manually, taxpayers receive tax returns that are already prefilled with data. Although tax returns prefilled by tax authorities should be highly trustworthy, errors in data input and data transmission can occur. For example, in the UK, "experts estimate that one in ten returns prefilled using HMRC's data will contain errors, which could lead to people paying too much or too little tax" (Telegraph, 2017). Moreover, innovative tools are using new techniques to produce prefilled tax returns

based merely on scanned bills and receipts. Nevertheless, techniques such as optical character recognition are still associated with errors.

While prefilling aims to help the taxpayer file a legally accurate tax return (OECD, 2017), mis-specified or incorrect numbers that are already included in a prefilled tax return might be associated with additional tax evasion. There is some evidence that the implementation of prefilled tax returns reduces tax compliance costs at the taxpayer's level (Goolsbee, 2006, Klun, 2009, Evans and Tran-Nam, 2010). However, there is almost no evidence regarding how prefilled – and particularly incorrectly prefilled – tax returns affect the compliance behavior of taxpayers. Slemrod (2016), for example, states in his recent literature review on tax compliance that "the direct effect of pre-population on compliance is unclear" (p. 64). Kleven et al. (2011) provide initial empirical results that prefilling can have compliance consequences. They find that tax compliance is higher for income that is subject to third-party reporting (already prefilled) than for self-reported income (no prefilling). In contrast, Kotakorpi and Laamanen (2016) and Gillitzer and Skov (2016) observe that partly prefilled tax returns have either no effect or even a slightly negative effect on reported taxable income.

In a controlled experiment, we investigate how prefilled tax returns affect compliance behavior by implementing and varying different prefilling scenarios simultaneously. We are able to clearly distinguish between correctly and incorrectly prefilled tax returns and can isolate their respective influences on tax compliance. Our set-up additionally allows us to study how individuals adjust prefilled tax returns and how adjustment behavior varies between different prefilling scenarios. In our lab experiment, we only vary how tax returns are prefilled, and we ensure that all monetary aspects such as tax rate, audit probability, and penalties are kept constant. Thus, monetary costs are not influenced by the prefilling of the tax returns, and we can exclude that prefilling changes tax compliance behavior through a change in the audit probability or penalty. This experimental design allows us to provide a clean test of how prefilling affects tax compliance behavior and provides a high level of internal validity. Consequently, we feel confident that conducting a laboratory experiment is an appropriate method of answering our research question.[7]

Our experiment consists of two parts. First, participants are given a real effort task to earn their pre-tax income. In the second part, participants have to file a tax return with six income

[7] Alm et al. (2015) report evidence that laboratory experiments in the field of tax compliance exhibit external validity. They show that behavioral patterns of subjects in the laboratory correspond to decision-making in naturally occurring settings. For a detailed discussion of the use of laboratory experiments for tax research issues, see also Alm (2010).

fields, one for each round of the real effort task. Participants have the opportunity to evade taxes: if they report a lower income than actually earned, they reduce their tax liability. However, there is a certain probability that participants will be audited. If a participant is caught evading, she is punished with a fine.

In our study, we use four treatments. In the first treatment, participants have to file a blank tax return (i.e., no income field is prefilled). In the other three treatments, participants are confronted with a prefilled tax return. Each participant can easily adjust the prefilled values to enter their own values. In the second treatment, each income field of the tax return is correctly prefilled. In the third and fourth treatments, we prefill randomly chosen income fields with an incorrect income. In the third treatment, the prefilled income is lower than the true income; in the fourth treatment, the prefilled income is higher than the true income. The remaining income fields are still correctly prefilled. This feature allows us to study our research questions not only in a between-subject design but also in a within-subject design.

Our main results are fourfold. First, we find that subjects are more tax compliant when the tax return is correctly prefilled compared to when the tax return is blank. This effect is meaningful and economically significant. In fact, we observe an increase of 15.5 percentage points in the average compliance level. Thus, correct prefilling enhances tax compliance. Second, if the income items are incorrectly prefilled with income below the true income, tax compliance is significantly lower than for correctly prefilled tax returns. The compliance level observed is the same as that for blank tax returns.

Third, the tax compliance level is higher for prefilled tax returns where the reported income is incorrectly high. The tax compliance level does not significantly differ from that of correctly prefilled tax returns. This finding implies an asymmetric effect of incorrect prefilling: If taxpayers would benefit from incorrect prefilling, prefilling has no effect on tax compliance level when compared with blank tax returns. In contrast, if taxpayers would suffer from incorrect prefilling, prefilling still has a positive effect on compliance compared to the case of blank tax returns, and compliance does not differ from compliance in the case of correctly prefilled tax returns.

Fourth, we find that individuals are aware of the incorrectly prefilled values and actively adjust them in their tax returns. Thus, we can rule out that a lack of awareness explains our findings. Moreover, as more than ninety percent of the incorrect values are adjusted by participants, we provide clear evidence that subjects do not stick to the default option and that the observed compliance behavior is not the result of a passive choice. Interestingly, we observe different adjustment behaviors depending on the type of prefilled values. Prefilled income

higher than true income is almost always adjusted downward. For prefilled income below true income, however, adjustments depend on the deviation from true income. If the deviation is small, subjects do not adjust prefilled values, but if the deviation is large, subjects adjust lower prefilled values upward.

Our results have several implications. Most importantly, we show that prefilled values and defaults have a meaningful influence on compliance behavior. This finding contributes to the general literature on cheating and lying behavior, where prefilling effects have not yet been studied (Gneezy, 2005, Ariely, 2012, Erat and Gneezy, 2012, Battigalli et al., 2013). However, as a consequence of the digital transformation, individuals and decision makers in organizations and institutions are often confronted with prefilled data and prefilled forms in almost all industries. In situations where dishonesty is of concern, prefilling and defaults might reduce dishonest behavior. For example, financial accounting and reporting software (e.g., SAP ERP) automatically generates several financial indicators on the performance of a manager's division, and those indicators are used for budget and investment decisions. Our results suggest that the presence of those defaults and the quality of those defaults both influence the willingness to misreport such indicators.

Our results further suggest that prefilling forms or data entry masks is preferable over non-prefilling. Providing correctly prefilled forms enhances compliance compared to the use of blank forms. Although the positive compliance effect induced by correctly prefilled forms vanishes in the case of beneficially incorrectly prefilled forms, compliance is still at the same level as we observed with blank forms. Furthermore, our results show that prefilled values are not revised if the prefilled values deviate only slightly from true values and the individual benefits from the incorrect prefilling. Consequently, due to errors that can occur by prefilling, we cannot expect to achieve the high compliance levels associated with correctly prefilled forms.

Furthermore, as prefilling tax returns is a service that the tax administration provides to taxpayers, our study is related to the service paradigm literature. For example, Alm et al. (2010a) show that services from the tax administration (e.g., agency-provided information) have a positive and significant impact on compliance behavior. Our paper also contributes to the slippery slope literature (Kirchler et al., 2008) and to the trust paradigm literature (Alm and Torgler, 2011). Both bodies of literature show that tax compliance is influenced not only by the power of the tax authority to conduct audits, collect taxes and punish tax evasion (enforcement paradigm) but also by the trust in authority. While an increase in the power of authority leads to higher enforced tax compliance, an increase in the trust in authority leads to higher voluntary

tax compliance (Wahl et al., 2010, Kastlunger et al., 2013, Kogler et al., 2013). If tax returns are correctly prefilled by the tax authority, the trust in authority might increase, and therefore tax compliance might increase as well. However, if tax returns are incorrectly prefilled, the trust in authority might be lower, leading to a decrease in tax compliance.

The remainder of this paper is structured as follows: In Section 2, we discuss the related literature and develop hypotheses. Section 3 describes the experimental design. We analyze compliance behavior in section 4 and adjustment behavior in section 5. Robustness tests are presented in section 6. Section 7 concludes.

2.2 Related Literature and Hypotheses

2.2.1 Prefilling and Tax Compliance

Our research questions of how prefilled – and particularly how incorrectly prefilled – tax returns influence tax compliance behavior are largely unexplored. Although some papers focus on prefilling (especially third-party reporting), the effect of incorrectly prefilled tax returns has not yet been studied. Kleven et al. (2011) analyze data from a tax enforcement experiment in Denmark and find that tax evasion is very low for income subject to third-party reporting and thus already prefilled in tax returns; however, they find that tax evasion is substantial for self-reported (i.e., not prefilled) income.

Kotakorpi and Laamanen (2016) use data from a natural experiment in Finland and examine tax reporting behavior when taxpayers receive prefilled tax returns. The authors observe that prefilling increases the number of deductions claimed but not the number of income items reported. Rather, the authors find a significant reduction in the number of reported items that were not prefilled. More importantly, on an aggregated level, they do not find that prefilled tax returns influence total taxable income or taxes paid.

Gillitzer and Skov (2016) use data from the Danish tax authority and examine the case of prefilled deductions. Contrary to their expectations, they find that the number of tax deductions claimed doubles and that the total value of deductions increases if tax-deductible charitable contributions are already prefilled in the tax return. The authors suggest that taxpayers neglect to claim their tax-deductible charitable contributions if they are not already prefilled.

Our study substantially differs from previous studies. First, we use a laboratory experiment that enables us to focus on the influence of prefilled tax returns on compliance behavior in a controlled environment. Second, whereas previous studies have had to make the simplifying assumption that items are correctly prefilled, we are able to clearly distinguish between

correctly and incorrectly prefilled tax returns and are able to isolate their respective influences on tax compliance. Third, previous studies have only analyzed the number of items claimed but not the actual compliance level. We design an experiment that enables us to analyze the level of tax compliance in more detail.

Fourth, our design allows us to control for the several potential explanations discussed by the studies mentioned above. Kleven et al. (2011), Gillitzer and Skov (2016) and Kotakorpi and Laamanen (2016) suggest that compliance is much higher for third-party reported (i.e., prefilled) items because the possibility of evading taxes is limited. We exclude this explanation with our experimental design, as our treatments offer the same opportunities for tax evasion in the cases of both prefilled and blank tax forms. Kotakorpi and Laamanen (2016) further discuss complexity effects as a possible explanation for changes in reporting behavior. We control for complexity by keeping the compliance decision in our experiment very simple. Participants have full information, there is no computation necessary, and complexity does not differ between our treatments.

2.2.2 Hypotheses

Initiated by the seminal papers of Becker (1968), Allingham and Sandmo (1972) and Yitzhaki (1974), a variety of papers have studied tax evasion and dishonest behavior.[8] The main monetary factors determining compliance level are the audit probability and penalty in case of disclosed non-compliance. More recently, studies have also been incorporating non-pecuniary factors such as social norms and moral/psychic costs of dishonest behavior (e.g., Kim, 2003, Fortin et al., 2007, Dulleck et al., 2016). In our study, the only variation is the prefilling in our treatments, and all monetary aspects are kept constant.[9] Prefilling has no influence on tax rate, audit probability, or penalty. Consequently, our experimental design ensures that monetary costs are not affected by different prefilling scenarios. Particularly, we can exclude that prefilling changes tax compliance behavior through a change in the audit probability or penalty. However, we argue that prefilling might influence compliance behavior through the following non-monetary phenomena observed in the literature.

---

[8] See Alm et al. (1995), Andreoni et al. (1998), Torgler (2002), Hofmann et al. (2008), Alm (2012), and Slemrod (2016) for excellent literature reviews.

[9] In accordance with the experimental tax compliance literature, we expect the tax compliance level to be greater than zero but below full compliance. We set the fiscal parameters (tax rate, audit probability and fine multiplier) in such a way that a purely payoff-maximizing (i.e., without moral concerns) and risk-neutral subject will always declare zero income (full evasion). Depending on the actual risk attitude, a purely payoff-maximizing and risk-averse subject will choose full evasion, full compliance or something in between.

First, if a tax return is prefilled with *correct* values, the default effect might have an impact on compliance. We define a correctly prefilled tax return as a prefilled tax return containing truthful values for all fields, thus resulting in an accurate tax base and tax liability for the taxpayer. Several studies find evidence for the default effect, which describes the preference of individuals to stay with a preset default option (in our case, prefilled income) rather than actively adjusting the default (Johnson and Goldstein, 2003, Mazar and Hawkings, 2015). If individuals stick to the default option and thus do not adjust correctly prefilled values, they will submit a legally accurate tax return and consequently behave in a tax compliant way. We therefore expect tax compliance to be higher in the case of correctly prefilled tax returns (i.e., prefilled income equals true income) compared to blank tax returns (i.e., tax returns are not prefilled).

Furthermore, tax compliance behavior might also be influenced by the anchoring effect if tax returns are already prefilled. According to Tversky and Kahneman (1974), the anchoring effect describes the influence of an initially presented value on decision making.[10] This influence can be caused by an insufficient adjustment of this value, which serves as a starting point, so that the final decision is assimilated toward this starting point (Tversky and Kahneman, 1974, Epley and Gilovich, 2001). Another suggestion is that the influence of an initially presented value is caused by selective accessibility, i.e., an unconscious activation of knowledge that is consistent with the presented anchor (Strack and Mussweiler, 1997, Chapman and Johnson, 1999). Regardless of the underlying mechanisms that account for the anchoring effect, one could expect that taxpayers completing returns are biased toward the starting values of the prefilled tax return. Thus, compared to the blank tax return scenario (without any starting value), one could expect tax compliance to be higher in the correctly prefilled tax return scenario.

Moreover, we argue that prefilled tax returns might influence the moral costs of tax evasion. The tax compliance literature suggests that higher moral costs increase the level of tax compliance (Erard and Feinstein 1994, Frey and Torgler, 2007, Dulleck et al., 2016). This literature argues, for example, that violating a social norm of honest and compliant behavior creates moral costs (such as anticipated shame and guilt) and that these moral costs become "a cost factor in evaluating one's likely advantages and disadvantages of tax evasion" (Kirchler, 2007, p. 64). Moreover, moral costs can arise through contradicting one's own intrinsic motivation to pay taxes or feelings of moral obligation (Frey, 1997, Frey and Torgler, 2007).

[10] For a detailed literature overview see Furnham and Boo (2011).

Moral costs associated with non-compliance reduce the utility of being non-compliant and therefore motivate compliance (Gordon, 1989, Bosco and Mittone, 1997, Fortin et al., 2007, Traxler, 2010, Blaufus et al., 2017). In our case of correct prefilling, tax evasion would require a deliberate adjustment of the prefilled values. We expect that the act of replacing correct values with incorrect numbers in order to evade taxes increases the moral costs of non-compliant behavior. Thus, with an increase in moral costs for tax evasive behavior, we expect that correctly prefilled tax returns lead to a higher tax compliance level compared to the case without prefilled income (blank tax returns).

As all explanations tend in the same direction, we formulate the following hypothesis:

Hypothesis 1: *The tax compliance level is higher when taxpayers are given correctly prefilled tax returns compared to blank tax returns.*

Second, we analyze how *incorrectly* prefilled tax returns that would result in tax savings for the taxpayers (i.e., prefilled income is lower than true income) affect tax compliance. If prefilled values are not adjusted, a taxpayer declares less income than is actually earned, resulting in a lower tax liability and thus resulting in tax evasion. If individuals have a preference to stay with a preset default (default effect) and therefore do not adjust the prefilled values, tax compliance will consequently be lower for lower prefilled than for correctly prefilled tax returns. If tax compliance behavior is influenced by anchoring effects, we would expect taxpayers to be assimilated toward the starting values of the prefilled tax returns. The prefilled lower income operates as an anchor toward lower income reporting. From this perspective, if tax returns are incorrectly prefilled with income values beneficial to the taxpayer, we again expect tax compliance to be lower compared to compliance when tax returns are correctly prefilled.

Moreover, we expect that incorrect prefilling has an opposite effect on moral costs than correctly prefilled tax returns. Research has shown that the moral evaluations of acts of omission and acts of commission differ significantly (omission bias). Individuals judge harmful omissions (e.g., failing to prevent harm) as less immoral than harmful commissions (e.g., actively imposing harm) (Spranca et al., 1991, Ritov and Baron, 1995, Cox et al., 2017). In the case of prefilled tax returns, failing to adjust incorrectly lower prefilled values is an act of omission, whereas a deliberate adjustment of correctly prefilled income is an act of commission. In accordance with the literature, we expect that individuals perceive the failure to adjust the

incorrect values as less immoral. Consequently, they will be less compliant in the case of lower prefilling than in the case of correct prefilling.

In addition, license to cheat and responsibility shifting may further reduce the moral costs of dishonest behavior in the case of incorrectly prefilled tax returns. As income is already incorrectly prefilled (e.g., similar to an incorrect data transmission by the tax authority), individuals might feel they have the license to behave in a tax-evasive way, or they may feel less responsible for submitting incorrect values. As the lower income value already entered in the return is "somebody else's mistake", an individual can evade taxes while maintaining a positive self-view (Mazar et al., 2008). Consequently, the moral costs of tax evasion are lower compared to the moral costs of tax evasion with correctly prefilled tax returns. Thus, we would expect a lower compliance level if prefilled tax returns are incorrect, particularly if the prefilled income is below the true income.

Overall, we formulate the following hypothesis:

Hypothesis 2: *Compared to the tax compliance level in the case of correctly prefilled tax returns, the tax compliance level is lower if taxpayers are given incorrectly prefilled tax returns that contain income below the true income.*

Third, we analyze how incorrectly prefilled tax returns that would result in tax disadvantages for the taxpayers (i.e., prefilled income is higher than true income) affect tax compliance. In particular, if prefilled values are not adjusted by a participant, she declares more income than she actually earned, and her corresponding tax liability is higher.

If taxpayers have a preference to stay with a preset default (default effect) and therefore do not adjust the prefilled values, we expect higher compliance levels on average in case of higher prefilling compared to correct prefilling. If we assume that behavior is affected by anchoring effects, we would expect taxpayers to be assimilated toward the prefilled starting values. In particular, the higher prefilled income works as an anchor toward higher income reporting. As a consequence, the anchoring effect should also lead to a higher tax compliance level for higher prefilled than for correctly prefilled tax returns. We formulate the following hypothesis:

Hypothesis 3a: *Compared to the tax compliance level in the case of correctly prefilled tax returns, the tax compliance level is higher if taxpayers are given incorrectly prefilled tax returns that contain income above the true income.*

In the case of higher prefilled tax returns, a downward adjustment from a higher prefilled value to the correct value is not associated with monetary costs because no taxes are evaded. Consequently, downward adjustments – at least to the correct value – are to be expected. However, an adjustment below the correct value is associated with costs. While the associated monetary costs are identical to those in the correctly prefilled scenario, the moral costs might be lower due to the incorrect prefilling. In particular, individuals might evaluate the adjustment of an incorrectly prefilled value as less immoral than the adjustment of a correct value. As the prefilled tax return already contains incorrect values, individuals might perceive this as a license to cheat. Moreover, individuals might feel unfairly treated by tax returns suggesting a higher tax than the correct tax. The literature shows that if taxpayers feel unfairly treated, their trust in and cooperation with tax authorities will be reduced, resulting in lower compliance levels (Hofmann et al. 2008, Kirchler et al. 2008). Consequently, from this perspective, the tax compliance level might be lower for higher prefilled tax returns than for correctly prefilled tax returns.

Hypothesis 3b: *Compared to the tax compliance level in the case of correctly prefilled tax returns, the tax compliance level is lower if taxpayers are given incorrectly prefilled tax returns that contain income above the true income.*

Table 1 provides an overview of the potential behavioral responses to different prefilling scenarios and the expected effects on tax compliance.

Table 1: Development of Hypotheses

	Hypothesis 1 (blank vs. correctly prefilled)		**Hypothesis 2** (correctly vs. lower prefilled)		**Hypothesis 3** (correctly vs. higher prefilled)	
	effect	**result**	**effect**	**result**	**effect**	**result**
default effect	accept correctly prefilled value	compliance increases	accept lower prefilled value	compliance decreases	accept higher prefilled value	compliance increases
anchoring effect	decision is assimilated toward correctly prefilled value	compliance increases	decision is assimilated toward lower prefilled value	compliance decreases	decision is assimilated toward higher prefilled value	compliance increases
moral costs	increase in moral costs	compliance increases	decrease in moral costs	compliance decreases	decrease in moral costs	compliance decreases
Total effect		compliance increases		compliance decreases		total effect unclear

Note: This table shows the influence of each discussed effect on tax compliance and the resulting total effect.

2.3 Experimental Design and Treatments

2.3.1 Experimental Design

We conduct a laboratory experiment consisting of two parts. The instructions are provided to the participants at the beginning of the experiment (see appendix A1). In the first part, participants work on a real effort task to earn their pre-tax income. In the second part, participants have to file a tax return to determine their tax liability.

We use a between-subject design with four treatments in which we vary whether and how participants' tax returns are prefilled (see section 3.2 for details). Subjects are randomly distributed to one of the four treatments. After participants have submitted their tax return, they are audited with a certain probability and have to pay a fine in case of a detected tax evasion. After the experiment, participants receive a tax-free show-up fee of EUR 4 as well as their earned pre-tax income minus their tax liability and minus a potential fine. Before the actual experiment is executed, we measure subjects' willingness to take risk with the Holt and Laury (2002) task (in EUR). The amount earned in the lottery task is also paid out to each participant at the end of the experiment.

In the first part of the experiment, we use the math puzzle task of Mazar et al. (2008). Participants see matrices with twelve numbers (each with two decimal places) on their screen and have to select the two numbers that add up to ten (e.g., 6.61 + 3.39 = 10). The math puzzle

is a search task in which participants have to put in some effort to correctly solve the puzzles to earn money. In each matrix, there are only two numbers that add up to ten. Participants play six rounds of the math puzzle task, each lasting three minutes, with a one-minute break between the rounds. In each round, they can solve a maximum of 20 puzzles. For every correctly solved math puzzle, a participant earns a pre-tax income of EUR 0.42 (EUR 0 otherwise). After each round, the participant's number of correctly solved math puzzles and the resulting earned income in that round are displayed. In order to be able to complete the tax return in the second part of the experiment, participants are requested to record their earned income after each round on a piece of paper at their workstation.

After finishing the real effort task, participants file a tax return in the second part of the experiment. A tax of 25% is levied on the declared income. Participants have the opportunity to evade taxes when declaring less income than actually earned. In the instructions, participants are explicitly asked to declare their earned income from part 1. Thus, when declaring less than they earned, participants engage in tax evasion. Unintentional tax evasion by the taxpayer is excluded by design, as our setting is very simple and participants are fully aware of their true income in each round. They do not have to perform any calculations; their only task is to declare the income in the six income fields of the tax return. For this purpose, they are asked to use the records they made on the piece of paper in the first part of the experiment. Nevertheless, if participants desire, they can press a button on the screen to have their actual earned income for each round displayed when they file their tax return.

There is a 30% probability that participants will be audited after they have submitted their tax returns.[11] If a participant is audited and her declared income is lower than her true income, she has to pay a fine that is twice the amount of the evaded taxes. This implies that in case of a detected tax evasion, the subject has to repay the evaded taxes plus additional penalty costs of 100% of the evaded taxes.[12] After subjects are informed about the audit outcome, the experiment is finished and the participants are asked to complete an ex-post questionnaire before they privately receive their payout in cash.

During the experiment, participants receive a "bank account", which is displayed to them after each part of the experiment. First, their total earned income in part 1 is credited to their account. Second, after participants have submitted their tax return, the resulting tax liability is

[11] While the probability of an actual audit of income tax returns by tax authorities is quite low in most countries, we chose an audit probability of 30% in order to take into account a possibly higher audit probability for prefilled income fields. This parameter is kept constant across all treatments.

[12] In cases where an audit reveals that the declared income is higher than the actual earned income, the subject gets the overpaid taxes back, but no additional costs occur.

debited to their account. If there is an audit and a tax evasion is detected, the resulting fine is debited to the account. As a final step, the show-up fee of EUR 4 and the payout from one randomly chosen lottery of the Holt and Laury (2002) task are added. The resulting balance is the participant's total payoff from the experiment.

Before the experiment starts, subjects have to successfully complete a comprehension test. They are asked several questions regarding the puzzle task, pre-tax income determination, tax liability determination, audit probability and payoff determination. The full set of questions can be found in our appendix A1.4.

2.3.2 Treatments

In the second part of the experiment, participants are given a tax return on which their income must be declared. The displayed tax return consists of six income fields, one for the income of each round of the real effort task. Below these fields, the resulting total declared income, as well as the resulting tax liability, is displayed to the participants. In our study, we use the following four treatments (between-subject design) in which we manipulate the tax return that participants have to file. Figure 1 presents our experimental design and highlights the treatment differences.

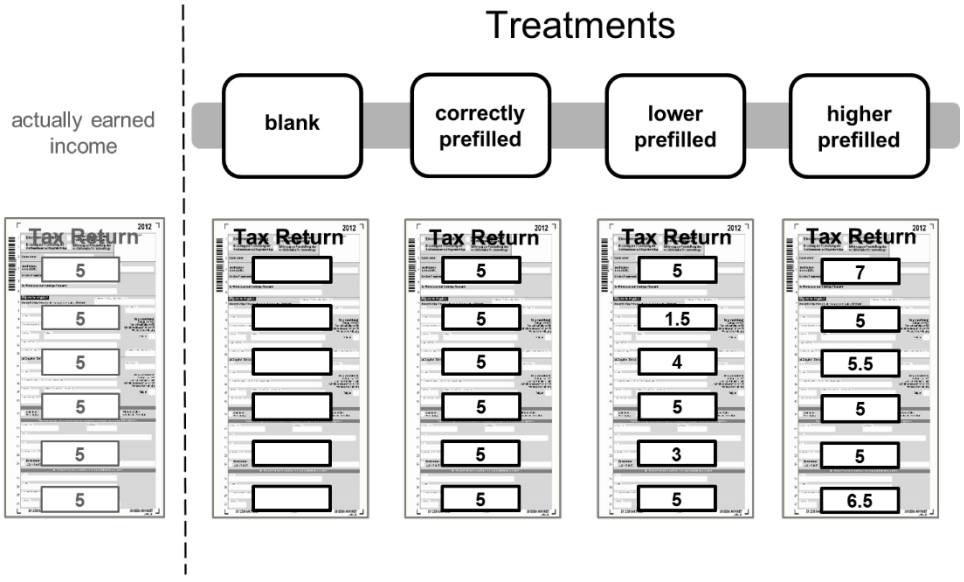


Figure 1: Experimental Design

Note: This figure highlights the differences among the treatments. Subjects are randomly distributed to one of the four treatments.

- *Treatment B (**b**lank)*: income fields are not prefilled, *blank* tax return.
- *Treatment CP (**c**orrectly **p**refilled)*: all income fields of the tax return are *correctly* prefilled (i.e., prefilled income equals true income).
- *Treatment LP (**l**ower **p**refilled)*: three income fields of the tax return are *incorrectly lower* prefilled to the advantage of the taxpayer (i.e., prefilled income is *lower* than true income); the other three income fields are correctly prefilled.
- *Treatment HP (**h**igher **p**refilled)*: three income fields of the tax return are *incorrectly higher* prefilled to the disadvantage of the taxpayer (i.e., prefilled income is *higher* than true income); the other three income fields are correctly prefilled.

In treatment B, participants have to file a blank tax return, meaning that no income field is prefilled. In the other three treatments, participants receive a tax return where all six income fields are already prefilled. All prefilled values can be adjusted by participants. In treatment CP, each income field of the tax return is correctly prefilled (i.e., contains participants' true income from each round of the real effort task). In both treatments LP and HP, we manipulate three randomly chosen income fields by prefilling an incorrect income. In treatment LP (HP), the incorrect income is lower (higher) than the corresponding actually earned income. The other three unmanipulated income fields are correctly prefilled. This within-subject variation enables us to analyze how the same subject responds to both an incorrect and correct prefilling.

For each incorrectly prefilled value, we randomly vary the level of deviation from the correct amount. Deviation is equally distributed and lies between 0% and 100% of the actually earned income. For treatment LP, this implies that the manipulated prefilled income lies between 0% (when deviation is 100%) and 100% (when deviation is 0%) of the true income. For treatment HP, this implies that the manipulated prefilled income lies between 100% (when deviation is 0%) and 200% (when deviation is 100%) of the true income. For example, suppose that the actually earned income is EUR 5. Consequently, the incorrectly prefilled income lies between EUR 0 and EUR 5 in treatment LP and between EUR 5 and EUR 10 in treatment HP. As deviation is determined randomly for each manipulated income field and thus naturally varies across these fields, this setting allows us to examine how different levels of deviation influence taxpayers' compliance behavior. Please note that for the manipulated income fields in treatments LP and HP, we ensure that deviation is above 0% (i.e., no deviation is not possible).

In the instructions, we ask participants to check whether the prefilled tax return is correct. Each prefilled value of the tax return can be changed as desired by the participant, with a

minimum of EUR 0 and a maximum of EUR 8.40 (20 correctly solved puzzles à EUR 0.42). The same applies to the values the participants are allowed to enter in the blank tax return in treatment B. Empty fields are not allowed. Moreover, participants can always use a computation button on the screen that displays the resulting total declared income and the resulting tax liability based on the inserted income values. There is no computation restriction, so participants can use the button as often as they want to.

We control for all other factors that may influence tax compliance, such as audit probability, fine, and tax rate, by keeping them constant between the different treatments. In our setting, the only variable manipulated is the prefilling of the tax returns. If there are differences in compliance, they can only result from our prefilling manipulation.

2.3.3 Sample and Data

The experiment was conducted at the computerized experimental laboratory of the University of Cologne (CLER) from August to October 2017. The experimental software was programmed and used with the software z-Tree (Fischbacher, 2007). Participants were recruited with ORSEE (Greiner, 2004). In total, 213 subjects (mainly undergraduate students, 122 females and 91 males) participated and earned, on average, EUR 19.04 in approximately 97 minutes (approximately EUR 11.78 per hour). A total of 42 subjects were randomly assigned to treatment B, 43 to treatment CP, 64 to treatment LP and 64 to treatment HP.[13] Over all treatments, we have 1,278 observations (213 subjects with 6 decisions per subject). Table 2 provides an overview on the main characteristics of our participants.

[13] In all treatments we have six observations per subject. As we manipulated only three of the six income fields in the treatments LP and HP (i.e., only three observations per subject for the manipulation), we decided to increase the chance of being assigned to these two treatments in order to increase the number of observations for the manipulated income fields.

Table 2: Main Characteristics of Participants

Variable	description	mean
female	female = 1; male = 0	57.28%
risk attitude	Holt & Laury (2002) risk measure; we use the number of high risk lottery (lottery B) choices per individual as our risk attitude measure	3.96 / 10
age	in years (19 to 69)	25
economics	study with more than one lecture in economics = 1 (0 otherwise)	43.66%
bachelor	education with a bachelor's degree = 1 (0 otherwise)	54.46%
tax experience	experience with tax returns = 1 (0 otherwise)	40.85%
tax knowledge	tax knowledge = 1 (0 otherwise)	12.68%
income	in Euro (monthly income after fixed costs)	370.35
tax morale	0 to 9; low tax morale = 0; high tax morale = 9	7.15
fairness	0 to 10; low perceived fairness of tax and control system in experiment = 0; high perceived fairness of tax and control system in experiment = 10	6.04
decision complexity	0 to 10; low perceived decision complexity in experiment = 0; high perceived decision complexity in experiment = 10	1.93
joy	0 to 10; felt no joy during experiment = 0; felt high joy during experiment = 10	4.40
anger	0 to 10; felt no anger during experiment = 0; felt high anger during experiment = 10	4.56
fear	0 to 10; felt no fear during experiment = 0; felt high fear during experiment = 10	1.92
guilt	0 to 10; felt no guilt during experiment = 0; felt high guilt during experiment = 10	1.48

Note: This table provides an overview of the individual characteristics of the 213 participants in our experiment.

2.4 Results: Declared Income

2.4.1 Descriptive Statistics and Nonparametric Statistics

Our measure for tax compliance is the declared share of true income. If a taxpayer declares the income she actually earned, she complies with the tax law. If she declares an income lower (higher) than the income actually earned, she acts in a tax evasive (over-compliant) way. To account for earned income heterogeneity, we use a relative measure and calculate the ratio between declared income and actually earned income for each single income field. As one subject declares income in six income fields, we have six observations per subject. For the following descriptive and nonparametric analyses, we calculated an average tax compliance level over these observations for each subject. For treatments LP and HP, this was performed

separately for manipulated and unmanipulated income fields. Figure 2 shows the mean values of declared share of income for each treatment.[14]

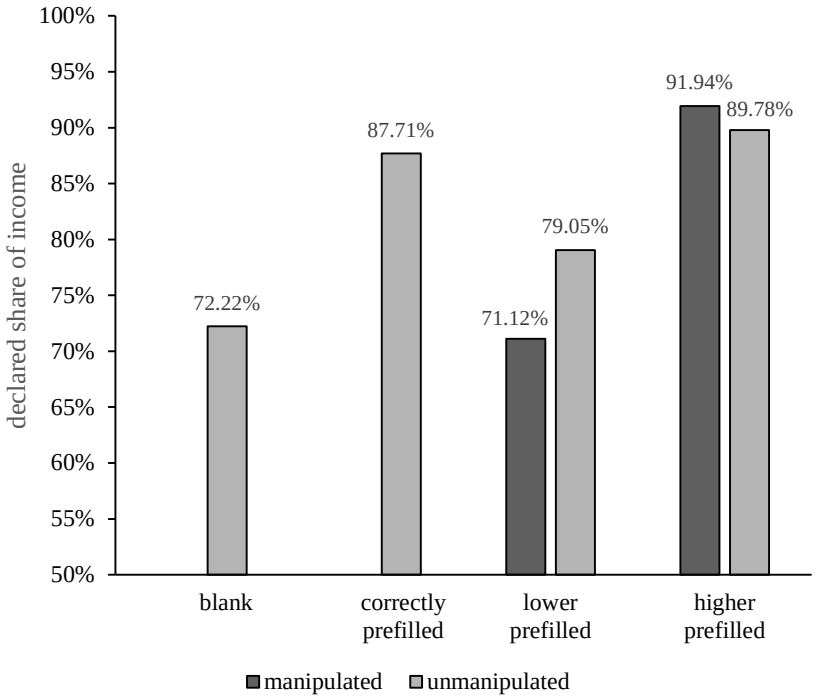


Figure 2: Declared Share of Income in our Treatments

Note: This figure shows the relative declared income for each treatment on average. For treatments LP and HP, the average declared income is shown separately for correctly and incorrectly prefilled income fields. As we have several observations per subject, we calculated an average tax compliance level for each subject.

The mean declared share of true income lies between 71.1% and 91.9% in our treatments. In line with the experimental tax compliance literature, average tax compliance is therefore higher than 0% but below 100% in all four of our treatments.

The mean declared share of true income in treatment blank (B) is 72.2%, and it is 87.7% in the correctly prefilled treatment (CP). In line with our first hypothesis, tax compliance significantly increases when tax returns are correctly prefilled compared to when blank tax returns are used (Mann-Whitney U-test, $p = 0.048$, two-tailed). Thus, hypothesis 1 is supported and we can formulate the following result:

[14] In our experiment, the largest deviation possible is 100% of the actually earned income. This implies that the manipulated prefilled values can range from 0% (in treatment LP) to 200% (in treatment HP) of the actually earned income. There is no reason for a declared income higher than 200% of the actually earned income. Therefore, we excluded observations (not subjects) with a declared income greater than 200% of the actually earned income. Please note that all stated results will stay the same if we include all observations in our analysis.

Result 1: *Compared to blank tax returns, the tax compliance level is significantly higher when tax returns are already correctly prefilled.*

The mean declared share of income in treatment LP is 71.1% for the incorrectly prefilled income fields. This share is significantly lower than the mean declared share of 87.7% in treatment CP (Mann-Whitney U-test, $p = 0.011$, two-tailed). Additionally, we can compare the declared share of income in a within-subject comparison, as three random income fields were incorrectly prefilled (manipulated) and the other three income fields were correctly prefilled (unmanipulated). For treatment LP, we find that the average declared share of income is significantly lower for the incorrectly prefilled income fields (71.1%) compared to correctly prefilled income fields (79.1%) (Wilcoxon signed-rank test, $p = 0.007$, two-tailed). Thus, hypothesis 2 is supported by both a between-subject (treatment CP vs. LP) and a within-subject (treatment LP) comparison:

Result 2: *Compared to correctly prefilled tax returns, the tax compliance level is significantly lower if tax returns are incorrectly prefilled with income values that are below the true income.*

In treatment HP, we observe a mean declared share of income of 91.9% for the incorrectly prefilled (manipulated) income fields. The difference between the declared share of income for the incorrectly prefilled income fields in treatment HP and correctly prefilled income fields in treatment CP is not statistically significant (Mann-Whitney U-test, $p = 0.738$, two-tailed). In the within-subject comparison we also find that the declared share of income for the incorrectly prefilled income fields (91.9%) does not statistically differ from the declared income for the correctly prefilled income fields (89.8%; Wilcoxon signed-rank test, $p = 0.795$, two-tailed). Thus, we do not observe a significant difference between incorrectly prefilled tax returns and correctly prefilled tax returns – neither in a between-subject comparison nor in a within-subject comparison. This result supports hypothesis 3a, whereas we cannot confirm hypothesis 3b.

Result 3: *The tax compliance level does not significantly differ when tax returns are incorrectly prefilled with a higher income compared to correctly prefilled tax returns.*

2.4.2 Regression Analyses

To corroborate our descriptive and nonparametric results, we also present a linear regression analysis. In contrast to the nonparametric tests, we now consider each observation of the six income fields per subject. The dependent variable is the declared share of true income of each subject in each income field. To account for dependency of the six decisions within one subject, we consider linear regression models with subject-specific random effects.[15] Moreover, we consider a vector of subject characteristics such as gender, age, and risk attitude.[16]

Our variables of interest are, however, indicator variables for the treatment variations. Treatment CP serves as the default. Thus, each coefficient of our treatment dummies measures the difference between the respective treatment and treatment CP.

Table 3 reports the regression results. In both models, all observations from treatments B and CP are considered. However, in specification (1), we consider only the incorrectly (manipulated) income fields, while in specification (2), we consider the correctly (unmanipulated) income fields of treatments LP and HP. Model 1 enables us to study whether prefilling manipulation has an effect on tax compliance level. Model 2 analyzes whether differences exist between treatments when income fields are not manipulated in the treatments LP and HP.

In both models, we observe a significant negative coefficient for treatment B. We find a lower tax compliance level in the treatment with blank tax returns compared to the treatment with correctly prefilled tax returns. Thus, we are able to confirm our result 1. Moreover, in model 1, we observe a significant negative coefficient for treatment LP. Hence, we find a lower tax compliance level compared to the treatment with correctly prefilled tax returns if prefilled income values are below true income. In contrast, we do not find a significant coefficient for treatment HP. Consequently, the tax compliance level does not statistically differ from that of correctly prefilled tax returns if the tax returns include prefilled income that is higher than true income. These findings confirm our results 2 and 3 in a between-subject comparison. Statistical significance between treatments B and LP, treatments B and HP, and treatments LP and HP was checked by Wald tests, and the resulting p-values are reported at the bottom of the table.

---

[15] We use linear regression models with random effects where the income field number is the time variable and the subject's identity number is the cross-sectional variable. Please note that we lose some observations/subjects in our regressions due to the restriction discussed in footnote 7.

[16] We use the number of high-risk lottery (lottery B) choices per individual from the Holt and Laury (2002) task as our risk attitude measure. The more risk-seeking an individual is, the higher is the number of her lottery B choices.

As a control group, we consider in specification (2) only the correctly prefilled income fields of treatments LP and HP. We find no compliance differences between subjects if income fields are correctly prefilled.

Table 3: Random Effects Linear Regressions, Between-Subject Comparison (dependent variable: declared share of true income)

	model 1	model 2
	manipulated	unmanipulated
blank (B)	-0.16**	-0.16**
	(0.08)	(0.08)
lower prefilled (LP)	-0.15**	-0.07
	(0.07)	(0.07)
higher prefilled (HP)	0.03	0.01
	(0.07)	(0.08)
individual controls	yes	yes
constant	0.72***	0.73***
	(0.21)	(0.21)
no. of observations	858	853
no. of subjects	211	210
R-squared		
within	0.0000	0.0000
between	0.3036	0.2640
overall	0.2635	0.2320
Wald test:		
B = LP	p = 0.9173	p = 0.2215
B = HP	p = 0.0089	p = 0.0196
LP = HP	p = 0.0054	p = 0.2129

Note: In this table, the results of random effects linear regressions are presented with the declared share of true income as dependent variable (regression coefficients, standard errors in parentheses). To analyze the influence of the prefilling manipulation in our experiment, we regress on dummy variables for the treatments B, LP and HP while treatment CP serves as the default. In both models, all observations from treatments B and CP are considered. However, in model 1, only the observations of the manipulated values of treatments LP and HP are included in the regression, while in model 2, only the observations of the unmanipulated values of treatments LP and HP are included. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 4 reports the outcome for our within-subject comparison. In model 3 (4) we analyze the differences within treatment LP (HP). In both models, the dummy variable "manipulated" takes the value of one if the declared share of income refers to a manipulated income field (0 otherwise). In model 3 (treatment LP), we observe a significant negative coefficient for the manipulation dummy on the declared share of income. Thus, the declared share of income is significantly lower for the lower prefilled income fields than for the correctly prefilled income fields. In contrast, we do not find a significant coefficient in model 4 (treatment HP). The declared share of income for the higher prefilled income fields does not differ from the correctly

prefilled income fields. As a result, we are also able to confirm our stated results 2 and 3 in a within-subject comparison by the regression analysis.

Table 4: Random Effects Linear Regressions, Within-Subject Comparison
(dependent variable: declared share of true income)

	model 3	model 4
	only treatment LP	only treatment HP
manipulated	-0.08***	0.02
	(0.02)	(0.02)
individual controls	yes	yes
constant	0.97**	0.67**
	(0.46)	(0.30)
no. of observations	380	373
no. of subjects	64	64
R-squared		
within	0.0564	0.0023
between	0.3845	0.4129
overall	0.3301	0.3318

Note: In this table, the results of random effects linear regressions are presented with the declared share of true income as dependent variable (regression coefficients, standard errors in parentheses). To analyze the differences within treatments, we use the dummy variable "manipulated", which takes the value 1 if the declared income refers to a manipulated income field (0 otherwise). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

In all models, we include individual specific variables such as gender, age, and risk attitude. In total, we incorporate the 15 individual variables reported in table 2 in our regressions. The individual variables are not displayed, but the complete set of all regression results is presented in our appendix A2 (tables A2.1 and A2.2). In line with the literature (e.g., Alm et al., 2010b, Rosenbaum et al., 2014), we observe that women are significantly more compliant than men, that subjects are less compliant the more risk-seeking they are, and that subjects with higher tax morale are more compliant. Furthermore, we find that the more complex the subjects perceive the tax-related decisions in our experiment to be, the more tax compliant they are.

We can summarize that our explorative and nonparametric findings are supported by our random effects linear regressions. In particular, we are able to confirm our finding that tax compliance is higher in the case of correctly prefilled tax returns compared to blank tax returns (result 1). Furthermore, we find support for the finding that tax compliance is lower in the case of incorrectly prefilled income that is below true income (result 2). Moreover, we find that tax

compliance does not differ when tax returns are incorrectly prefilled with income values that are higher than true income (result 3).

2.4.3 Treatment Blank vs. Different Prefilled Variations

We further analyze the differences between the treatment with blank tax returns and each of the treatments with prefilled tax returns. Table 5 reports the results of linear regression models with random effects for a between-subject comparison, as used in section 4.2. In this case, the non-prefilled treatment B serves as the default. This analysis allows us to compare each prefilled scenario with the case of blank tax forms. Again, we consider observations from either the manipulated (model 5) or unmanipulated (model 6) income fields of treatments LP and HP.

Table 5: Random Effects Linear Regressions, Between-Subject Comparison (dependent variable: declared share of true income)

	model 5	model 6
	manipulated	unmanipulated
correctly prefilled (CP)	0.16**	0.16**
	(0.08)	(0.08)
lower prefilled (LP)	0.01	0.09
	(0.07)	(0.07)
higher prefilled (HP)	0.19***	0.17**
	(0.07)	(0.07)
individual controls	yes	yes
constant	0.57***	0.57***
	(0.20)	(0.20)
no. of observations	858	853
no. of subjects	211	210
R-squared		
within	0.0000	0.0000
between	0.3036	0.2640
overall	0.2635	0.2320

Note: In this table, the results of random effects linear regressions are presented with the declared share of true income as dependent variable (regression coefficients, standard errors in parentheses). To analyze the influence of the prefilling manipulation in our experiment, we regress on dummy variables for the treatments CP, LP and HP, while treatment B serves as the default. In both models, all observations from treatments B and CP are considered. However, in model 1, only the observations of the manipulated values of treatments LP and HP are included in the regression, while in model 2, only the observations of the unmanipulated values of treatments LP and HP are included. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

The results again suggest that correctly prefilled tax returns significantly increase compliance compared to blank tax returns. When comparing blank tax returns with incorrectly prefilled tax returns, we find no significant difference in compliance if the prefilled income is below the true income. Thus, the positive compliance effect induced by correctly prefilled tax returns vanishes if the prefilled tax return contains errors that lead to tax savings for the taxpayer. If the tax return is prefilled with income values that are higher than the true income, tax compliance is significantly higher than in the case of non-prefilled tax returns. As the difference between treatment CP and treatment HP is statistically insignificant (see table 3 and section 4.2), errors in the tax return that lead to tax disadvantages for the taxpayer do not further increase compliance compared with correctly prefilled tax returns.

2.4.4 Influence of Tax Morale on Treatment Effects

We analyze how our treatment effects are influenced by the individual tax morale of our subjects. In an ex-post questionnaire, we use an adapted question from the World Values Survey that is widely applied in the tax compliance literature (e.g., Slemrod 2003, Alm and Torgler, 2006). We asked: "How do you evaluate the following statement?: Cheating on tax if you have the chance..." Answers were given on a 10-point Likert scale from "…is always justifiable" = 0 to "…is never justifiable" = 9. The observed median (mean) tax morale level in our sample is 8 (7.15).

We use a median split to categorize our subjects into two subsamples. Subjects with a tax morale level of 8 and higher are categorized in the "high tax morale" subsample. Subjects with a level below the median are categorized in the "low tax morale" subsample. Figure 3 shows the mean declared share of income in all treatments for both subsamples. For the subsample with high tax morale (Panel A), we observe the same results (including statistical significance) as reported for the whole sample. For the subsample with low tax morale (Panel B), however, we do not observe any significant differences between our treatments – neither for the between-subject nor for the within-subject comparisons. Therefore, we can conclude that prefilling has a strong impact on compliance behavior for individuals with high tax morale but not for individuals with low tax morale.

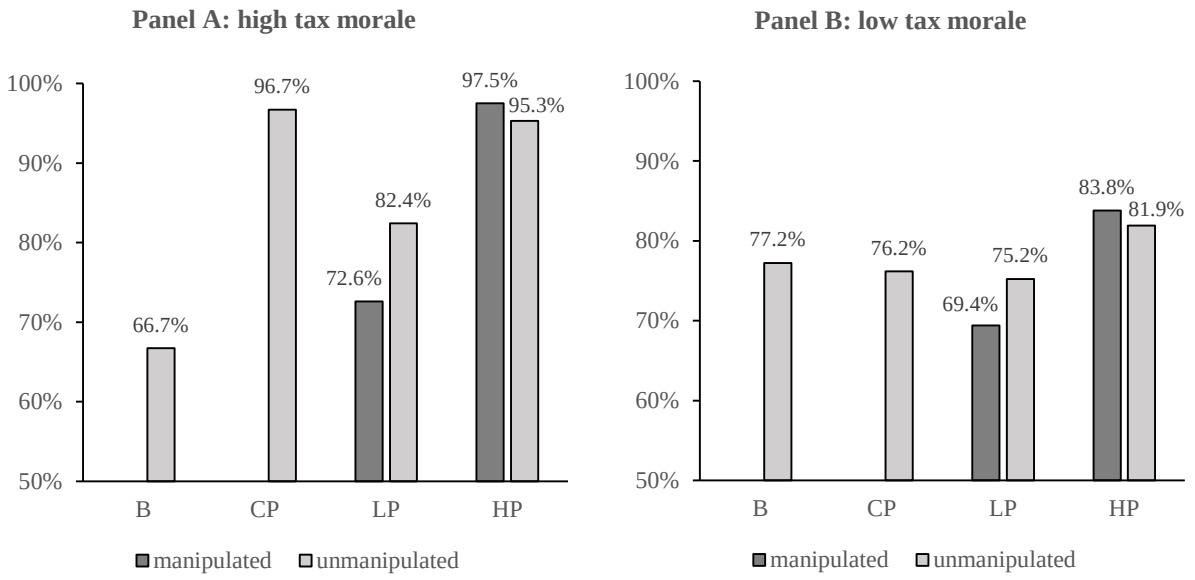


Figure 3: Declared Share of Income for Low and High Tax Morale Subsamples

Note: This figure shows the relative declared income for each treatment, on average, separately for subjects with high tax morale (Panel A) and subjects with low tax morale (Panel B). For treatments LP and HP, the average declared income is shown separately for correctly and incorrectly prefilled income fields. As we have several observations per subject, we calculated an average tax compliance level for each subject.

2.5 Results: Adjustments of Prefilled Values

In this section, we analyze whether and, if so, to what extent prefilled tax returns are adjusted. Adjustments are defined by the difference between declared and prefilled income. To account for income heterogeneity, we again use a relative measure and normalize to actually earned income.[17] Participants' upward (downward) adjustment of prefilled income of +10% (-10%) indicates that the declared share of income is increased (reduced) by 10% of the actually earned income. If, for example, the earned income is 10 and the prefilled value 6, the declared share of income is 7 (5) in case of a 10% upward (downward) adjustment. Table 6 shows the

[17] We normalize to actually earned income because this allows a clear comparison between the adjustments across treatments. An alternative would be to normalize to the prefilled value (i.e., difference between declared and prefilled values divided by prefilled value). However, this makes a direct comparison across treatments difficult. For example, suppose that earned income is 10 and prefilled values are 6 in treatments LP and 14 in HP (i.e., prefilled values deviate by -40% and +40%, respectively). If an adjustment is made by 1 in both treatments, the alternative measure will be +/-16.67% for treatment LP and +/-7.14% for HP. As prefilled values are (by design) always lower in LP than in HP, the measure will always be higher in LP, although adjustments are identical in absolute terms. With our measure, this issue is avoided and reveals adjustments of +/-10% for both treatments.

distribution and mean values of the adjustments made in our treatments for all observations, as well as for the correctly and incorrectly prefilled income fields separately.

Table 6: Difference between Declared and Prefilled Income (Adjustments)

		correctly prefilled			incorrectly prefilled		all observations		
adjustment		CP	LP	HP	LP	HP	CP	LP	HP
downward	%	23.5%	33.7%	25.5%	26.8%	92.1%	23.5%	30.3%	59.3%
	mean	-0.80	-0.69	-0.59	-0.43	-0.66	-0.80	-0.57	-0.64
no adjustment	%	**70.4%**	**63.7%**	**69.6%**	**6.3%**	**6.9%**	70.4%	35.0%	37.8%
	mean	0	0	0	0	0	0	0	0
upward	%	6.1%	2.6%	4.9%	66.8%	1.1%	6.1%	34.7%	2.9%
	mean	0.61	0.68	0.66	0.52	0.24	0.61	0.52	0.58
total	%	100%	100%	100%	100%	100%	100%	100%	100%
	mean	-0.15	-0.21	-0.12	0.23	-0.60	-0.15	0.01	-0.36

Note: This table shows the distribution and mean values of the adjustments of prefilled values for all observations as well as for correctly and incorrectly prefilled income fields separately. Adjustments are defined as the difference between declared and prefilled income (normalized).

Most importantly, we observe that almost all incorrectly prefilled income fields are adjusted (approx. 93%). Only 6.3% (6.9%) of incorrectly prefilled values are accepted by the subjects as prefilled in treatment LP (HP). In contrast, 64% to 70% of all correctly prefilled values are accepted and not adjusted. This gives clear evidence that subjects are aware of the incorrectly prefilled values and that they do not remain passive or stick to the prefilled values.

Result 4: *Subjects are aware of incorrectly prefilled income fields and adjust them.*

With respect to adjustment differences across our scenarios, we observe a similar decision pattern in each treatment for correctly prefilled income fields. In all three prefilled treatments, 64% to 70% of these income fields are unadjusted, 24% to 34% are adjusted downward, and only a negligible minority (3% to 6%) are adjusted upward. The mean adjustments are negative in all treatments and lie between -0.12 and -0.21.

In contrast, crucial differences are observed for the incorrectly prefilled income fields. Whereas almost all incorrectly prefilled income fields (92%) are adjusted downward in treatment HP, only 27% are adjusted downward in treatment LP. In the latter, the majority

(67%) is adjusted upward. This leads to a positive mean adjustment level of +0.23 in treatment LP and a negative level of -0.60 in HP.[18]

Result 5: *Whereas in treatment HP, almost all incorrectly prefilled income fields are adjusted downward, in treatment LP, the majority of income fields are adjusted upward.*

Figure 4 shows the mean adjustments for the incorrectly prefilled income fields in treatments LP and HP. This figure also exhibits the difference between prefilled and earned income and between declared and earned income (each normalized to actually earned income).

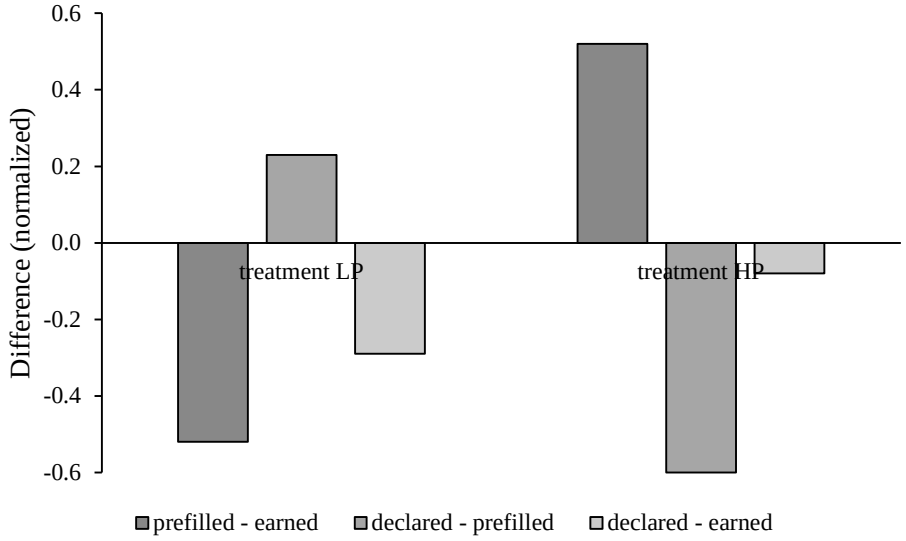


Figure 4: Adjustments in Case of Incorrectly Prefilled Tax Returns

Note: This figure shows the mean adjustments for the incorrectly prefilled income fields in treatments LP and HP (declared - prefilled). Adjustments are defined as the difference between declared and prefilled income (normalized). The figure also presents the difference between prefilled and earned income and between declared and earned income (each normalized to actually earned income).

As prefilled values deviate from correct income randomly between 0% and 100% of actually earned income (equal distribution), the normalized difference between prefilled and earned income is, on average, approximately -0.5 in treatment LP and +0.5 in treatment HP. In treatment LP, we find that the observed mean adjustments for the incorrectly prefilled income

[18] Interestingly, in treatment LP, the positive mean adjustment level of +0.23 in the case of incorrectly prefilled values is approximately as high as the negative adjustment level of -0.21 in the case of correctly prefilled values. Thus, adjustments cancel each other out on an aggregated level (-0.01). In contrast, subjects on average adjust prefilled income downward in cases of correctly (-0.12) and incorrectly (-0.60) prefilled values in treatment HP. Consequently, the mean adjustment level is negative (-0.36) on an aggregated level.

fields (+0.23) do not compensate the prefilling error (-0.52). Thus, a negative (normalized) difference between declared and true income results (-0.29). In treatment HP, the observed mean adjustments (-0.60) overcompensate the prefilling error (+0.52). Accordingly, we observe a negative (normalized) difference between declared and true income (-0.08). However, overcompensation is not large enough to achieve the low tax compliance level observed in treatment LP (-0.29 vs. -0.08). Consequently, declared share of income is ultimately lower in treatment LP than in HP.

Result 6: *In treatment LP, adjustments do not compensate lower prefilling. In treatment HP, adjustments overcompensate higher prefilling.*

In a final step, we analyze how the level of deviation of prefilled income from true income influences adjustment behavior. Figure 5 shows the adjustments made by subjects depending on the deviation of prefilled income from actually earned income. In treatment HP, we observe an almost linear relationship between adjustments and deviations. Consequently, prefilled values are adjusted in accordance with the deviation level: the larger the prefilled income deviates from true income, the larger the downward adjustment. On average, the difference between declared and prefilled income equals -0.35 for small deviation levels (deviations below or equal to 50%) and -0.80 for large deviation levels (deviations above 50%). In both cases, adjustments differ significantly from zero (two-tailed, one sample median test, p-value < 0.001).

In treatment LP, we observe no linear relationship. For small deviation levels, we find almost no adjustments. If adjustments are made, they are statistically insignificant (two-tailed, one sample median test, $p = 0.545$). In contrast, for large deviation levels, we find nearly constant adjustments in a range of +0.4 to +0.6. The mean difference between declared and prefilled income equals +0.51 and differs significantly from zero (p-value < 0.001). Consequently, if the prefilled income deviates only slightly from the true income, subjects keep the lower prefilled value without adjusting it. Above a deviation of 50%, subjects adjust the prefilled income upwards.

Result 7: *Subjects make no adjustments and keep lower prefilled income values if the prefilled values do not differ too greatly from the true values. For large deviations from true income, however, subjects adjust lower prefilled income values upward. In contrast, higher prefilled income values are almost always adjusted downward.*

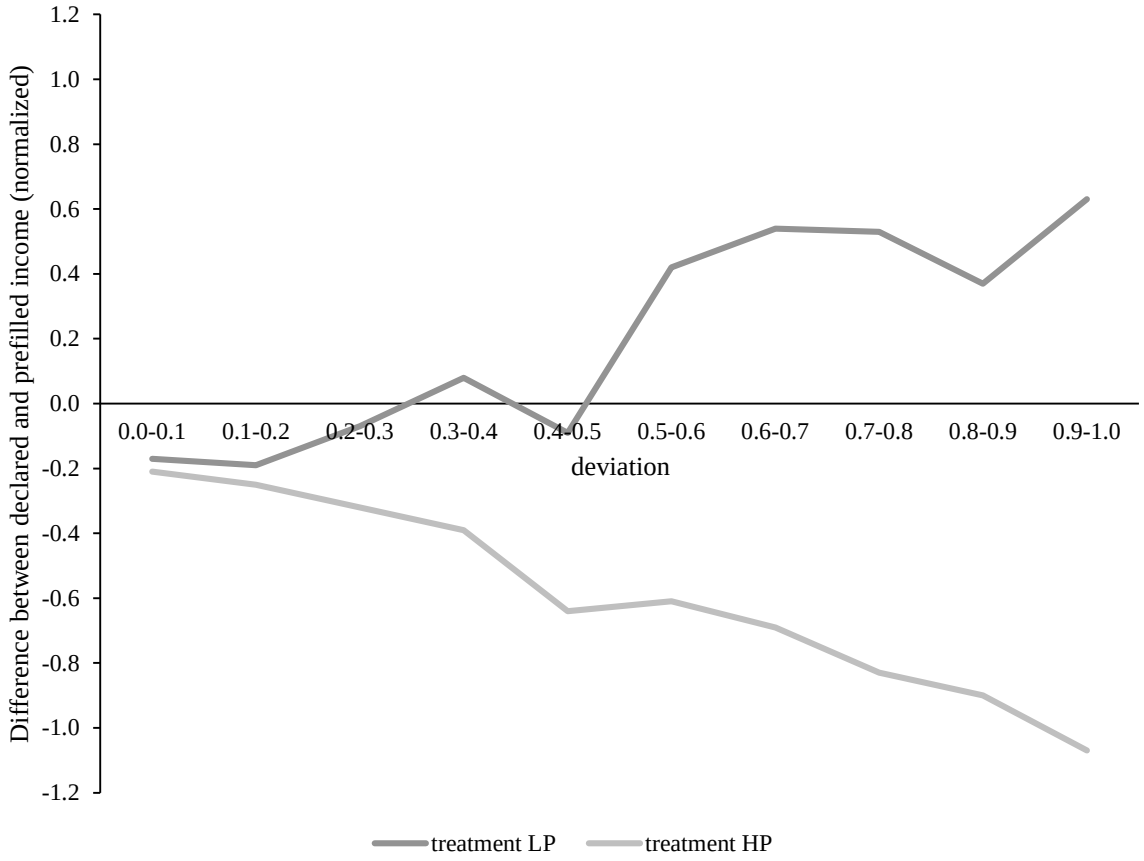


Figure 5: Deviation and Adjustments

Note: This figure shows subjects' adjustments dependent on the deviation of prefilled income from true income for treatments LP and HP. Adjustments are defined as the difference between declared and prefilled income (normalized).

As a robustness test, we reran our regression analyses separately for small (below or equal to 50%) and large (above 50%) deviations. The regression results are reported in tables A2.3 and A2.4 in our appendix A2. For both subsamples, we observe the same pattern as before, and all our stated results are robust to this variation. Consequently, whereas adjustments depend on the deviation level, the relative declared share of income does not.

2.6 Robustness Tests

Perceived Audit Probability

There is ample evidence that an increased audit probability increases tax compliance (see Torgler, 2002, for an overview). Although our prefilling variation has no influence on the objective audit probability, prefilling might increase the subjective probability of an audit. If the tax return is prefilled by the tax authority, a taxpayer might believe that changing the prefilled values will increase the probability that the tax authority will conduct an audit. To exclude this as an explanation, we first checked whether participants were aware of the audit

probability of 30%. For this purpose, participants had to state the audit probability before (in our comprehension test) and after the experiment (in our ex-post questionnaire). Only one of the 213 participants gave an incorrect answer on the ex-post questionnaire.[19] Second, and more importantly, we asked our participants in our ex-post questionnaire: "How did you perceive the audit probability in the experiment?" (10-point Likert scale from "very low" to "very high"). Over all treatments, the mean answer was 3.6. Differences across treatments were small and statistically insignificant (Kruskal-Wallis test, $p = 0.291$, two-tailed).[20] Consequently, we were able to exclude that the compliance differences across treatments were driven by different levels of perceived audit probability.

Individual Variables across Treatments

We also checked whether the distribution of individual characteristics (such as gender, risk attitude, tax morale, and age) differs across treatments. For each of our 15 individual variables reported in table 2, we either applied the chi-squared test (for binominal variables) or Kruskal-Wallis test (for ordinal and interval variables). We did not observe significant differences in our treatments with respect to individual variables (all p-values above 0.1, two-tailed). The only exception occurred with respect to fairness (Kruskal-Wallis test, $p = 0.036$, two-tailed). In particular, in our ex-post questionnaire, we asked: "How fair did you perceive the tax and control system that was applied in the experiment to be?" (11-point Likert scale from "very unfair" to "very fair"). Mean values across treatments were 6.38 in treatment B, 6.67 in CP, 6.02 in LP and 5.41 in HP. The pairwise comparison revealed significant differences between treatments B and HP as well as between treatments CP and HP (Mann-Whitney U-test, both p-values below 0.05, two-tailed). Consequently, the tax regime was perceived as significantly less fair in treatment HP, where the tax returns were incorrectly prefilled to the disadvantage of the subjects.[21]

Influence of Earned Income on Tax Compliance

In the tax compliance literature, there is much evidence that tax compliance decreases with income (see, for example, Grundmann and Lambsdorff, 2017, for a recent literature review and recent results). In line with this finding, we observed a significantly negative correlation

[19] That incorrect answer was 10% and was given in treatment HP.

[20] Mean values across treatments were 3.43 in treatment B, 3.26 in CP, 3.95 in LP and 3.61 in HP.

[21] The ex-post questionnaire was completed after all decisions were made and after the information was provided regarding whether the subject's tax declaration was audited. As the number of audits did not differ significantly across treatments (chi squared test, $p = 0.459$, two-tailed), the lower perceived fairness observed in treatment HP cannot be attributed to a higher audit frequency in this treatment.

between earned income (in EUR) and (relative) declared share of true income. Over all treatments, Spearman's rank correlation coefficient equaled -0.23 and was highly significant (p-value < 0.01, two-tailed).[22] We reran our regression analyses and added earned income as a control variable (not reported). All of our stated results were robust to this variation. In all regressions, the coefficient of earned income was significantly negative (all p-values below 0.01, two-tailed).

Consistent Tax Compliance Behavior

As is common in tax compliance experiments, we observed that often participants consistently chose either full compliance or full evasion. Over all treatments, 48.8% of our subjects declared the actual income truthfully in each of the six income fields (always full compliance). In contrast, 12.7% of our participants chose to always declare zero income (always full evasion). Consequently, 61.5% of all subjects revealed consistent tax compliance behavior over the six compliance decisions. See table 7 for treatment details. Although the distributions did not differ significantly across treatments (two-tailed, chi-squared test, all p-values above 0.3), we reran our regression analyses and excluded subjects who revealed such consistent behavior. Regression results are reported in table A2.5 and A2.6 in our appendix A2. As expected, treatment effect sizes increased. Nevertheless, all of our stated results are robust to this variation.

Table 7: Consistent Tax Compliance Behavior

	all treatments	treatment B	treatment CP	treatment LP	treatment HP
always full compliance	48.8%	45.2%	58.1%	42.2%	51.6%
always full evasion	12.7%	16.7%	11.6%	15.6%	7.8%
always full compliance or full evasion	61.5%	61.9%	69.8%	57.8%	59.4%

Note: This table presents an overview of the share of subjects that reveal consistent tax compliance behavior in our experiment. Consistent means that the subjects chose either full compliance or full evasion for all six income fields in their tax return.

Order Effects

In our experiment, the six different income fields were arranged one below the other in the tax return. In a robustness test, we checked whether compliance levels were influenced by order

[22] Spearman's rank correlation coefficient was between -0.19 and -0.26 in each treatment and was highly significant (all p-values below 0.01, two-tailed).

effects. It might be, for example, that individuals evaded more (or less) in the first than in the last income field. Although we randomly varied which three income fields were manipulated in treatments LP and HP, we wanted to make sure that order effects did not bias our previous analyses. We tested each treatment separately, and in treatments LP and HP, we considered manipulated and unmanipulated income fields separately. We did not find any significant differences in the compliance levels among the six income fields (Kruskal-Wallis test, all p-values above 0.6, two-tailed). Thus, an individual's compliance did not depend on the position of the income field in the tax return, and order effects can be excluded.

2.7 Conclusion

As a consequence of the digital transformation, the process of completing forms has fundamentally changed. In this study, we conduct a controlled experiment with 213 participants to analyze how prefilled forms influence compliance behavior. We frame our experiment as filing the annual income tax return.

We observe that tax compliance is significantly higher when tax returns are correctly prefilled compared to when tax returns are not prefilled (i.e., blank tax returns). Consequently, correct prefilling enhances tax compliance. More importantly, we study the influence of incorrectly prefilled tax returns and find asymmetric results. On the one hand, if prefilled income is lower than true income, we find no positive effect of prefilling: compared to the setting with blank tax returns, tax compliance is on the same level. However, compared to the setting with correctly prefilled tax returns, compliance is significantly lower when prefilled income is lower than true income. On the other hand, if prefilled income is higher than true income, prefilling still has a positive effect: compliance is on the same level as with correctly prefilled tax returns and is therefore significantly higher than with blank tax returns.

One merit of our experimental design is that it enables us to examine how individuals adjust prefilled income. We find clear evidence that individuals actively make adjustments. In particular, more than 93% of the incorrectly prefilled values are changed by participants. This result enables us to exclude lack of awareness (that prefilled values are incorrect) as a potential explanation. Moreover, the adjustment behavior clearly shows that individuals do not have a tendency to remain passive, and compliance behavior is therefore not the result of a passive choice. Participants do not stick to the default option (default effect) but rather make use of their opportunity to change the incorrectly prefilled values.

When analyzing the adjustment behavior for incorrect income values, we observe different adjustment behaviors depending on the direction of error. We find that higher prefilled income is nearly always adjusted downward. In contrast, adjustments of lower prefilled income depend on how much the prefilled income deviates from the true income. If prefilled values deviate only slightly from true income, individuals make no adjustments and keep the lower prefilled values. For larger deviations, however, subjects adjust lower prefilled values upward. We also observe that, on average, these adjustments do not compensate the initial lower prefilling.

Our study suggests that prefilling can help to increase compliance but that the quality of prefilling matters. In cases where the tax authority is able to prefill tax returns with data from employers, social insurance agencies or banks, the likelihood of an incorrect prefilling might be rather small. However, prefilled values will generally not match with actual values if an electronic tax return program carries forward the values entered in the tax forms of previous years. Our results suggest that particularly in cases of small deviations, prefilled values are not revised if taxpayers benefit from the incorrect prefilling. Additionally, a recent trend is the use of software applications that automatically generate prefilled tax returns after scanning pay slips, bills and receipts (e.g., taxbutler in Germany). However, incorrectly prefilled tax returns caused by technical errors of optical character recognition could lead to serious tax compliance issues, as reported in our study.

Our results also provide an answer to the following question: Should we make use of prefilled forms or prefilled entry masks? Indeed, prefilled forms reduce compliance costs. However, what about compliance behavior on an aggregated level? We observe that correctly and higher prefilled tax returns significantly increase tax compliance compared to blank tax returns. This finding supports the utilization of prefilled forms. By contrast, the positive compliance effect induced by correctly prefilled tax returns vanishes if the prefilled tax return contains errors that lead to tax savings for the taxpayer. However, tax compliance for lower prefilled tax returns is, on average, at the same level as for blank tax returns. From this perspective, our results suggest that prefilling does not always outperform non-prefilling in every situation but that prefilling is preferable over non-prefilling on an aggregated level.[23]

Furthermore, our results indicate that observed prefilling effects are mainly driven by individuals with high tax morale. In this case, prefilling has a strong impact on compliance behavior. For individuals with relatively low tax morale, however, the positive effects of

[23] Nevertheless, further research is needed for an analysis of long-term effects. If taxpayers, for example, are repeatedly confronted with incorrectly higher prefilled tax returns, it is plausible that trust in authority erodes over time, resulting in lower compliance levels (Kirchler et al., 2008, Alm and Torgler, 2011).

prefilling compared to non-prefilling should not be expected. This finding also implies that the influence of prefilling on compliance behavior might differ across countries. Alm and Torgler (2006), for example, report higher tax morale for the United States and Switzerland than for Spain, Germany, UK, or France. They argue that a direct democracy (e.g., more political participation possibilities) and high institutional quality might increase tax morale.

Our results contribute to the general literature on cheating and lying behavior (Gneezy, 2005, Ariely, 2012). We provide evidence that prefilled forms and defaults significantly influence compliance. Thus far, this effect is neither identified in the general lying and cheating literature nor identified as a potential explanation for how prefilling influences decision-making in the tax compliance literature.

Our study does have some limitations. First, we assume complete information about true income, while in reality, taxpayers might sometimes have incomplete information about their true income. Uncertainty might amplify the tendency for taxpayers to rely on prefilled values. In this case, the prefilled value implies an informative signal. Relying more on prefilled values increases tax compliance in the case of correctly and higher prefilled tax returns but decreases tax compliance in the case of lower prefilled tax returns. Consequently, the compliance gaps between lower and correctly prefilled tax returns, as well as the gaps between lower and higher prefilled tax returns, should increase further. However, the resulting difference between the correctly and higher prefilled tax return scenarios is not obvious.

Another issue is how information is distributed. In our paper, we assume symmetric information between the taxpayer and the tax authority. In reality, information about the taxpayer's income is often asymmetrically distributed, and the taxpayer can exploit this asymmetry strategically. If prefilled values reflect the information set of the tax authority, we might therefore observe that tax compliance further decreases in the lower prefilled treatment. However, we might not observe significant changes in the higher prefilled treatment because a taxpayer can submit evidence (e.g., receipts, invoices) to clarify why she has not accepted the prefilled values. Further research is also required here.

In our study, we ensured that all monetary aspects were kept constant to examine the behavioral impacts of prefilling on compliance behavior in a clean setting. In many real-life situations, however, prefilling might also change the monetary dimension. For example, if tax returns are prefilled with data already known by the tax authority (e.g., in the case of third-party reporting), changing these values is likely to increase the probability that the tax authority conducts an audit. Furthermore, tax evasion might be easier to convict and punish if a taxpayer

intentionally adjusts correctly prefilled tax returns downward. Again, further research is needed on how these effects add to our identified prefilling effects.

One limitation of our study is that we use students as subjects. Although this has several strong advantages (e.g., homogenous sample, high cognitive capability, low opportunity costs to ensure incentive compatibility), our results have to be treated with caution regarding external validity. However, as we are not interested in how subjects solve complex case studies where expertise is crucial or where a special context is important, we decided to use students. There is much evidence that student decision-making does not differ significantly from that of professionals and non-students – especially if the complexity of the applied experimental task is low like it is in our experiment (Alm et al., 2015, Depositario et al., 2009, Remus, 1996, Ashton and Kramer, 1980, Elliot et al., 2007). Therefore, we feel confident that using students as subjects is appropriate in our setting.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323–338.

Alm, J. (2010). Testing behavioral public economics theories in the laboratory. *National Tax Journal*, 63(4), 635–658.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1), 54–77.

Alm, J., Bloomquist, K. M. and McKee, M. (2015). On the external validity of laboratory tax compliance experiments. *Economic Inquiry*, 53(2), 1170–1186.

Alm, J., Cherry, T., Jones, M. and McKee, M. (2010a). Taxpayer information assistance services and tax compliance behavior. *Journal of Economic Psychology*, 31(4), 577–586.

Alm, J., Martinez-Vazquez, J. and Torgler, B. (Eds.) (2010b). Developing alternative frameworks for explaining tax compliance. Routledge.

Alm, J., Sanchez, I. and De Juan, A. (1995). Economic and noneconomic factors in tax compliance. *Kyklos*, 48, 3–18.

Alm, J. and Torgler, B. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology*, 27, 224–246.

Alm, J. and Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635–651.

Andreoni, J., Erard, B. and Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, 36, 818–860.

Ariely, D. (2012). The honest truth about dishonesty: How we lie to everyone – especially ourselves. Harper, New York.

Ashton, R. H. and Kramer, S. S. (1980). Students as surrogates in behavioral accounting research: Some Evidence. *Journal of Accounting Research*, 18, 1–13.

Battigalli, P., Charness, G. and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, 93, 227–232.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 169–217.

Blaufus, K., Bob, J., Otto, P. E. and Wolf, N. (2017). The effect of tax privacy on tax compliance – An experimental investigation. *European Accounting Review*, 26(3), 561–580.

Bosco, L. and Mittone, L. (1997). Tax evasion and moral constraints: some experimental evidence. *Kyklos*, 50(3), 297–324.

Chapman, G. B. and Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79, 115–153.

Cox, J. C., Servátka, M. and Vadovič, R. (2017). Status quo effects in fairness games: reciprocal responses to acts of commission versus acts of omission. *Experimental Economics*, 20(1), 1–18.

Depositario, D., Nayga, R., Wu, X. and Laude, T. P. (2009). Should students be used as subjects in experimental auctions? *Economics Letters*, 102, 122–124.

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M. and Torgler, B. (2016). Tax compliance and psychic costs: behavioral experimental evidence using a physiological marker. *Journal of Public Economics*, 134, 9-18.

Elliot, W., Hodge, F., Kennedy, J. and Pronk, M. (2007). Are MBA students a good proxy for nonprofessional Investors? *The Accounting Review*, 82, 139–168.

Epley, N. and Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological science*, 12(5), 391–396.

Erard, B. and Feinstein, J. (1994). The role of moral sentiments and audit perceptions in tax compliance. *Public Finance*, 49, 70–89.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58, 723–733.

Evans, C. and Tran-Nam, B. (2010). Managing tax system complexity: building bridges through prefilled tax returns. *Australian Tax Forum*, 25, 245−274.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.

Fortin, B., Lacroix, G. and Villeval, M. C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11-12), 2089–2112.

Frey, B. S. (1997). Not Just for the Money. An Economic Theory of Personal Motivation. Edward Elgar Publishing, Cheltenham, UK.

Frey, B. S. and Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of comparative economics*, 35(1), 136–159.

Furnham, A. and Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, 40, 35–42.

Gillitzer, C. and Skov, P. E. (2016). The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark. EPRU Working Paper Series.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 9, 384–394.

Goolsbee, A. (2006). The simple return: Reducing America's tax burden through return-free filing. Brookings Institution Hamilton Project Discussion Paper 2006-04.

Gordon, J. P. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European economic review*, 33(4), 797–805.

Greiner, B. (2004). An online recruitment system for economic experiments. *Forschung und wissenschaftliches Rechnen GWDG Bericht 63*. edited by K. Kremer and V. Macho. Göttingen, Gesellschaft für Wissenschaftliche Datenverarbeitung, 79–93.

Grundmann, S. and Lambsdorff, J. (2017). How income and tax rates provoke cheating – An experimental investigation of tax morale. *Journal of Economic Psychology*, 63, 27–942.

Hofmann, E., Hoelzl, E. and Kirchler, E. (2008). Preconditions of voluntary tax compliance: Knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. *Journal of Psychology*, 216, 209–217.

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

Johnson, E. J. and Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339.

Kastlunger, B., Lozza, E., Kirchler, E. and Schabmann, A. (2013). Powerful authorities and trusting citizens: The Slippery Slope Framework and tax compliance in Italy. *Journal of Economic Psychology*, 34, 36–45.

Kim, Y. (2003). Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. *Journal of Public Economics*, 87(7-8), 1591–1616.

Kirchler, E. (2007). The economic psychology of tax behaviour. New York: Cambridge University Press.

Kirchler, E., Hoelzl, E. and Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, 29, 210–225.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.

Klun, M. (2009). Prefilled income tax returns: Reducing compliance costs for personal income taxpayers in Slovenia. *Financial Theory and Practice*, 33(2), 219–233.

Kogler, C., Batrancea, L., Nichita, A., Pantya, J., Belianin, A. and Kirchler, E. (2013). Trust and power as determinants of tax compliance: Testing the assumptions of the slippery slope framework in Austria, Romania and Russia. *Journal of Economic Psychology*, 34, 169–180.

Kotakorpi, K. and Laamanen, J. P. (2016). Prefilled income tax returns and tax compliance: Evidence from a natural experiment. Tampere Economic Working Paper 104.

Mazar, N., Amir, O. and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644.

Mazar, N. and Hawkins, S. A. (2015). Choice architecture in conflicts of interest: Defaults as physical and psychological barriers to (dis)honesty. *Journal of Experimental Social Psychology*, 59, 113–117.

OECD (2017). Third-party data management – the journey from post-assessment crosschecking to pre-filling and no-return approaches. In: Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies. OECD Publishing.

Remus, W. (1996). Will behavioral research on managerial decision making generalize to managers? *Managerial and Decision Economics*, 17, 93–101.

Ritov, I. and Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes*, 64(2), 119–127.

Rosenbaum, S. M., Billinger, S. and Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45, 181-196.

Slemrod, J. (2003). Trust in public finance. *Public Finance and Public Policy in the New Century*, ed. S. Cnossen and H.-W. Sinn, 49–88. Cambridge, USA: The MIT Press.

Slemrod, J. (2016). Tax compliance and enforcement: New research and its policy implications. Ross School of Business Working Paper No. 1302.

Spranca, M., Minsk, E. and Baron, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, 27(1), 76–105.

Strack, F. and Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of personality and social psychology*, 73(3), 437–446.

Telegraph (2017). Taxpayers face chaos as HMRC replaces traditional tax returns with pre-filled forms. 26th September 2017, available online: http://www.telegraph.co.uk/news/2017/09/26/taxpayers-face-chaos-hmrc-replaces-traditional-tax-returns-pre/.

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

Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, 26(1), 89–103.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124–1131.

Wahl, I., Kastlunger, B. and Kirchler, E. (2010). Trust in authorities and power to enforce tax compliance. An empirical analysis of the "Slippery Slope Framework". *Law & Policy*, 32, 383–406.

Yitzhaki, S. (1974). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3(2), 201–202.

Appendix

A1 Instructions

Appendix A1 includes the translated instructions (from German). All participants received the general instructions in print. Before the actual experiment was executed, subjects participated in the Holt and Laury (2002) task. The instructions for this task (first experiment) were displayed on the computer screen. After that, participants received the specific instructions for the actual (second) experiment in print.

A1.1 General Instructions

Thank you for participating in this experimental study. For your participation you receive a participation fee of 4 Euros.

The study consists of two experiments in which you have the possibility to earn money and a questionnaire at the end of the study. How much money you earn depends on your decisions and on chance. The instructions explain how you can influence how much money you earn in this study by your decisions.

It is important that you understand the instructions. Please do not hesitate to ask questions. If you have a question, please raise your hand. We will come to you to answer your question. Please do not ask your question loudly. You can write on the instructions or set markers. Please do not take the instructions home, but give them back to us at the end of the study.

The analysis of the experiment will be anonymous. We will never link your name with the data generated in the experiment. You will not learn the identity of any other participant, neither before nor after the experiment. Also the other participants will not learn your identity. At the end of the experiment, you have to sign a receipt to confirm the payments you received. This receipt will only be used for accounting purposes.

We would like to inform you that you are not allowed to communicate with other participants or leave your seat throughout the whole experiment. Please switch off your mobile phone and put it in your bag.

The calculator, the pen and the sheet of paper (for notes) that are lying on your desk, can be used.

At the end of the study you will receive your payout privately and in cash. Your total payout consists of your payout of the first experiment, plus your payout of the second experiment and the participation fee.

The instructions for the first experiment will be displayed on your computer screen.

A1.2 Instructions for the Holt and Laury (2002) Task

Please choose one of the two lotteries A or B in each of the following 10 decision situations.

You will make a decision for all 10 situations, but your payout from the first experiment is determined only by the one situation that is randomly drawn by the computer after the second experiment.

In each situation, you can either earn 2.00 € or 1.60 € from lottery A and either 3.85 € or 0.10 € from lottery B. The probabilities of winning, however, vary from situation to situation. The further down you move in the table, the higher is the probability of the higher payment and the lower is the probability of the lower payment.

After the first experiment and the second experiment are completed, the computer randomly draws (with the same probability) one of the 10 decision situations. After that, the computer determines your payout from the lottery that you have chosen in this decision situation by a second random draw. For that, the computer uses the probabilities for the higher payment and the lower payment according to the chosen decision situation.

decision	Lottery A	Your decision		Lottery B
		A	B	
1.	2.00 € with 10% or 1.60 € with 90%	o	o	3.85 € with 10% or 0.10 € with 90%
2.	2.00 € with 20% or 1.60 € with 80%	o	o	3.85 € with 20% or 0.10 € with 80%
3.	2.00 € with 30% or 1.60 € with 70%	o	o	3.85 € with 30% or 0.10 € with 70%
4.	2.00 € with 40% or 1.60 € with 60%	o	o	3.85 € with 40% or 0.10 € with 60%
5.	2.00 € with 50% or 1.60 € with 50%	o	o	3.85 € with 50% or 0.10 € with 50%
6.	2.00 € with 60% or 1.60 € with 40%	o	o	3.85 € with 60% or 0.10 € with 40%
7.	2.00 € with 70% or 1.60 € with 30%	o	o	3.85 € with 70% or 0.10 € with 30%
8.	2.00 € with 80% or 1.60 € with 20%	o	o	3.85 € with 80% or 0.10 € with 20%
9.	2.00 € with 90% or 1.60 € with 10%	o	o	3.85 € with 90% or 0.10 € with 10%
10.	2.00 € with 100% or 1.60 € with 0%	o	o	3.85 € with 100% or 0.10 € with 0%

A1.3 Instructions for Main Experiment

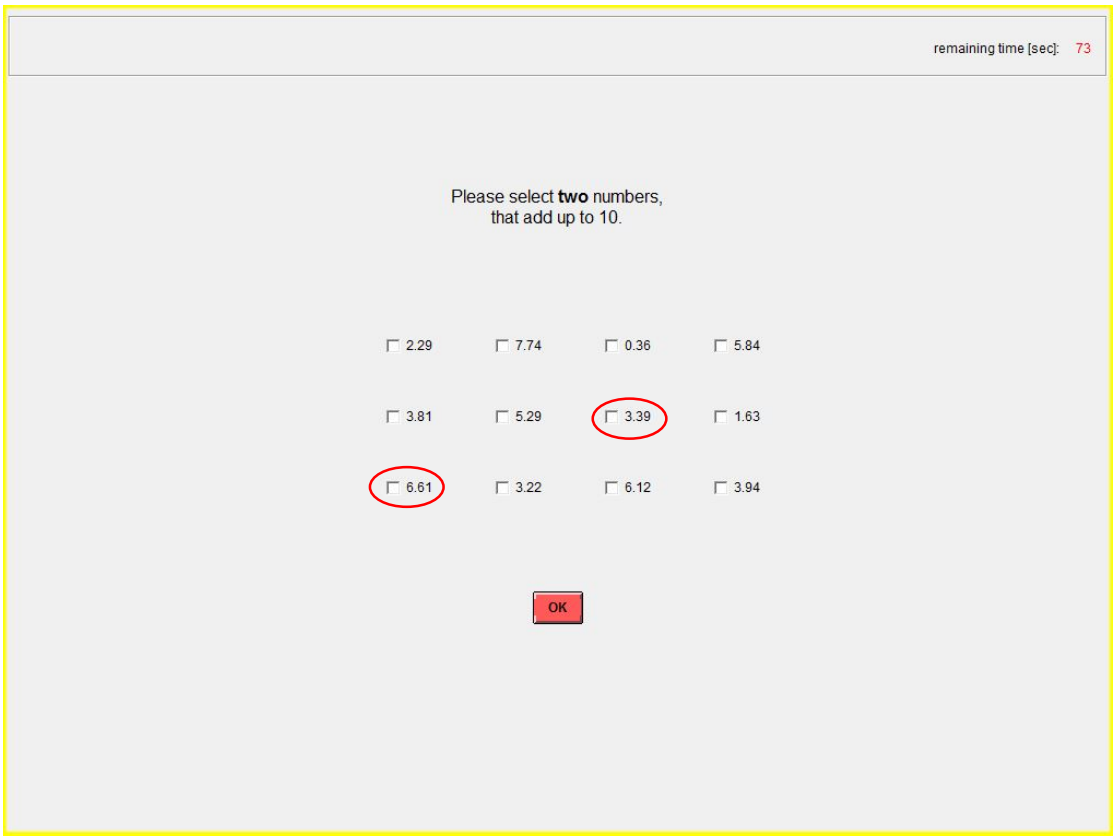
The second experiment consists of two parts. In part 1 of the experiment you have the possibility to earn money. In part 2 of the experiment you complete a tax return over your income.

Part 1

Puzzle Task

The first part of the experiment consists of a math puzzle task. The screen of your computer will display a puzzle with a matrix of 12 numbers. The numbers have two decimal places. These two numbers, that add up to 10, should be found in the matrix. In each matrix with 12 numbers, there are only 2 numbers that exactly add up to 10.

The screen with the matrix looks as follows:



Part 1 of the experiment consists of 6 rounds and in each round you have 3 minutes time to solve as many math puzzles as you like. A maximum of 20 puzzles can be solved in each round. After each round you have a one-minute break before the next round starts.

Your earned income depends on the number of puzzles you solved correctly. For each correctly solved puzzle, you receive an income of 0.42 €. For each incorrectly solved puzzle, you receive an income of 0 €. If you do not solve any puzzle correctly in one round, you earn 0 € in that round. If you solve every puzzle correctly in one round, you earn 8.40 € in that round. After each round, your number of correctly solved puzzles and the resulting income earned in that round will be displayed to you. Your income in each round therefore is as follows:

Earned income in each round =
number of correctly solved puzzles in the corresponding round x 0.42 €

Your total earned income results when adding up the earned income of all six rounds.

After you have earned your pre-tax income (= total earned income) in part 1 of the experiment, you have to complete a tax return on your income in the second part of the experiment. In order to be able to complete the tax return, please note down your earned income after each round on the provided sheet of paper at your place. The sheet of paper serves to help you completing the tax return. Please do not give the sheet of paper back to us, but take it home or dispose of the paper.

Part 2

Tax return

In this study, you have to complete a fictional tax return on your earned income, in order to determine a fictional tax. This means, in part 2 you shall declare the income you earned in part 1. There will be a tax of 25% on the declared income that was stated by you in the tax return. The tax revenue will be used inter alia to finance future research projects at the University of Cologne.

[Treatment B (blank)

After you have finished part 1 of the experiment, your tax return will be displayed on your screen. The tax return consists of 6 rows, one for the income of each round. At the end of each row you find a field in which you enter the income you want to declare. Your declared income can be lower, equal to, or higher than your actually earned income.

After you have filled out the 6 income fields, you can press the button "Submit tax return", in order to submit your tax return. After submission, you cannot modify your tax return anymore.

Before submitting your tax return, you can also press the button "Compute tax liability" in order to get the total declared income and the resulting tax liability on the basis of your provided information in the tax return displayed. If you want to modify your data, press the button "Modify tax return". You can then change your data in the six income fields. You can press "Compute tax liability" and "Modify tax return" until you are ready with your tax return. As soon as you want to submit your tax return, press the button "Submit tax return".]

[Treatment CP, LP and HP (prefilled)

After you have finished part 1 of the experiment, your tax return will be displayed on your screen. The tax return consists of 6 rows, one for the income of each round. The income of each round is already prefilled by the computer. Additionally, you get the total declared income and the resulting tax liability displayed. The prefilled values can deviate from your actually earned income. Please check the prefilled values in the tax return. Your declared income can be lower, equal to, or higher than your actually earned income.

To submit your tax return, please press the button "Submit tax return". After submission, you cannot modify your tax return anymore.

If you want to modify the data in your tax return, press the button "Modify tax return". You can then change the values in the six income fields. If you want to get the new resulting total declared income and the resulting tax liability displayed, press the button "Compute tax liability". You can press "Compute tax liability" and "Modify tax return" until you are ready with your tax return. As soon as you want to submit your tax return, press the button "Submit tax return".]

The tax payable amounts to 25% of your total declared income:

Tax = 0.25 x total declared income

Audit of the tax return

With a probability of 30% your tax return is audited. If you are audited and the declared income of a round does not coincide with your actually earned income of the corresponding round, you have to repay the undeclared tax. Additionally, a fine is charged at the same amount.

Tax repayment = sum of undeclared taxes

Fine = sum of undeclared taxes

The undeclared tax for the income of each round is:

Undeclared tax = 0.25 x (earned income – declared income)

If you are audited and you declared a higher income than you actually earned in that round, you get a tax refund of the overpaid tax. There will be no fine in that case.

Your personal payout of the 2. Experiment

During the 2. Experiment you will receive a personal "bank account", on which your payout relevant amounts will be posted during the experiment. Your account balance will be displayed to you after each part of the experiment. After part 1, your total earned income of the six rounds will be posted to your account and displayed to you. After you have submitted your tax return, the resulting tax burden of the tax return is posted to your account. If there is an audit and the declared income for each round does not coincide with the actually earned income in the corresponding round, the tax repayment and the resulting fine is debited to the account. If you receive a tax refund, it will be posted to your account.

Your account balance at the end of the 2. experiment is your personal payout of the 2. experiment. Your payout is calculated as follows:

Payout from the 2. experiment = *total earned income in part 1*
- *tax liability from part 2*
- *possible tax repayment*
- *possible fine*
+ *possible tax refund*

Before the 2. experiment begins, you are asked to answer some questions at your computer. The answering of the questions serves to verify your understanding and is not payoff relevant.

A1.4 Comprehension Test

In our comprehension test, subjects have to correctly answer the following questions:

1. How many experiments does the study include?
   Possible answers:
   - *1*
   - *2*
   - *3*
   - *4*
   - *More than 4*

2. How is the earned income for each round determined?
   Possible answers:
   - *The earned income for each round is determined by multiplying the number of correctly solved puzzles in that round with 0.42 €.*
   - *The earned income for each round is determined by multiplying the total number of correctly solved puzzles in all rounds with 0.42 €.*
   - *The earned income for each round is determined by multiplying the number of correctly solved puzzles in that round with 0.24 €.*

3. When is a puzzle solved correctly?
   Possible answers:
   - *When the number 10 is chosen.*
   - *When so many numbers are chosen, that they sum up to ten.*
   - *When those two numbers are chosen, that add up to ten.*

4. Which of the following statements regarding the computation of the tax liability is correct?
   Possible answers:
   - *The tax amounts to 25% of the total declared income.*
   - *The tax amounts to 25% of the total earned income.*

5. With what probability (in percent) will there be an audit of the tax return?
   Possible answer: between 0 and 100%

6. How is your personal payout for the second experiment determined when there is no audit?
   Possible answers:
   - *Payout = total earned income in part 1*
   - *Payout = total declared income in part 2 – tax liability in part 2*
   - *Payout = fix payout of 4 €*
   - *Payout = total earned income in part 1 – tax liability in part 2*

A2 Regressions

Table A2.1: Random Effects Linear Regressions, Between-Subject Comparison (dependent variable: declared share of income)

	model 1	model 2
	manipulated	unmanipulated
blank (B)	-0.16**	-0.16**
	(0.08)	(0.08)
lower prefilled (LP)	-0.15**	-0.07
	(0.07)	(0.07)
higher prefilled (HP)	0.03	0.01
	(0.07)	(0.08)
risk attitude	-0.05***	-0.06***
	(0.02)	(0.02)
female	0.21***	0.17***
	(0.05)	(0.05)
age	0.01	0.01
	(0.00)	(0.00)
economics	-0.09*	-0.08
	(0.05)	(0.05)
bachelor	0.09	0.11*
	(0.05)	(0.06)
tax experience	-0.08	-0.08
	(0.05)	(0.06)
tax knowledge	-0.10	-0.07
	(0.08)	(0.08)
tax morality	0.02*	0.03*
	(0.01)	(0.01)
fairness	-0.01	-0.00
	(0.01)	(0.01)
decision complexity	0.05***	0.05***
	(0.01)	(0.01)
joy	-0.01	-0.00
	(0.01)	(0.01)
anger	-0.02*	-0.02*
	(0.01)	(0.01)
fear	-0.01	-0.01
	(0.01)	(0.01)
guilt	-0.00	0.00
	(0.01)	(0.01)
income	0.00	0.00
	(0.00)	(0.00)
constant	0.72***	0.73***
	(0.21)	(0.21)
no. of observations	858	853
no. of subjects	211	210
R-squared		
within	0.0000	0.0000
between	0.3036	0.2640
overall	0.2635	0.2320
Wald test:		
B = LP	p = 0.9173	p = 0.2215
B = HP	p = 0.0089	p = 0.0196
LP = HP	p = 0.0054	p = 0.2129

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.2: Random Effects Linear Regressions, Within-Subject Comparison (dependent variable: declared share of income)

	model 3	model4
	only treatment LP	only treatment HP
manipulated	-0.08***	0.02
	(0.02)	(0.02)
risk attitude	-0.08**	-0.07**
	(0.04)	(0.03)
female	0.28***	0.17**
	(0.10)	(0.08)
age	-0.01	0.00
	(0.01)	(0.01)
economics	-0.10	0.07
	(0.11)	(0.09)
bachelor	0.13	0.08
	(0.11)	(0.09)
tax experience	-0.04	-0.01
	(0.10)	(0.09)
tax knowledge	0.03	-0.10
	(0.16)	(0.12)
tax morality	0.03	0.04*
	(0.02)	(0.02)
fairness	-0.03	0.00
	(0.02)	(0.02)
decision complexity	0.09***	0.06***
	(0.03)	(0.02)
joy	0.01	-0.01
	(0.02)	(0.02)
anger	-0.01	-0.02
	(0.02)	(0.01)
fear	-0.02	-0.01
	(0.03)	(0.02)
guilt	-0.02	0.02
	(0.03)	(0.02)
income	0.00	0.00
	(0.00)	(0.00)
constant	0.97**	0.67**
	(0.46)	(0.30)
no. of observations	380	373
no. of subjects	64	64
R-squared		
within	0.0564	0.0023
between	0.3845	0.4129
overall	0.3301	0.3318

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.3: Random Effects Linear Regressions, Separate for Small and Large Deviations, Between-Subject Comparison (dependent variable: declared share of income)

	model 1	model 2
	manipulated	manipulated
	deviation <= 0.5	deviation > 0.5
blank (B)	-0.15**	-0.15*
	(0.07)	(0.09)
lower prefilled (LP)	-0.20***	-0.16*
	(0.07)	(0.08)
higher prefilled (HP)	0.04	0.07
	(0.07)	(0.08)
individual controls	yes	yes
constant	0.77***	0.72***
	(0.20)	(0.23)
no. of observations	653	684
no. of subjects	192	197
R-squared		
within	0.0000	0.0000
between	0.3618	0.2642
overall	0.2870	0.2228
Wald test:		
B = LP	p = 0.5329	p = 0.9399
B = HP	p = 0.0059	p = 0.0057
LP = HP	p = 0.0005	p = 0.0025

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.4: Random Effects Linear Regressions, Separate for Small and Large Deviations, Within-Subject Comparison (dependent variable: declared share of income)

	model 3		model 4	
	only treatment LP	only treatment LP	only treatment HP	only treatment HP
	deviation <= 0.5	deviation > 0.5	deviation <= 0.5	deviation > 0.5
manipulated	-0.09***	-0.08***	0.003	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
individual controls	yes	yes	yes	yes
constant	1.01**	0.96**	0.66**	0.69**
	(0.44)	(0.47)	(0.32)	(0.32)
no. of observations	280	290	268	289
no. of subjects	64	64	64	64
R-squared				
within	0.0596	0.0495	0.0000	0.0063
between	0.4121	0.3646	0.3913	0.4078
overall	0.3631	0.3150	0.3433	0.3108

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.5: Random Effects Linear Regressions without Subjects Revealing a Consistent Tax Compliance Behavior, Between-Subject Comparison
(dependent variable: declared share of income)

	model 1	model 2
	manipulated	unmanipulated
blank (B)	-0.30**	-0.25*
	(0.14)	(0.15)
lower prefilled (LP)	-0.29**	-0.08
	(0.12)	(0.13)
higher prefilled (HP)	-0.13	-0.16
	(0.13)	(0.14)
individual controls	yes	yes
constant	0.73**	0.63*
	(0.33)	(0.37)
no. of observations	305	300
no. of subjects	80	79
R-squared		
within	0.0000	0.0000
between	0.5379	0.3840
overall	0.4163	0.2847
Wald test:		
B = LP	p = 0.9165	p = 0.1461
B = HP	p = 0.1103	p = 0.4177
LP = HP	p = 0.1023	p = 0.4739

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.6: Random Effects Linear Regressions without Subjects Revealing a Consistent Tax Compliance Behavior, Within-Subject Comparison
(dependent variable: declared share of income)

	model 3	model4
	only treatment LP	only treatment HP
manipulated	-0.19***	0.04
	(0.04)	(0.05)
individual controls	yes	yes
constant	-0.22***	2.38***
	(1.34)	(0.81)
no. of observations	158	145
no. of subjects	27	26
R-squared		
within	0.1361	0.0061
between	0.5282	0.8265
overall	0.3290	0.4892

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Chapter 3

Partial Prefilling: Compliance Behavior and Adjustment Behavior for Prefilled Values

Partial Prefilling: Compliance Behavior and Adjustment Behavior for Prefilled Values

Nadja Müller

University of Cologne

Abstract

Following the trend of digital transformation, an increasing number of tax administrations offer prefilled forms in the tax declaration process, aiming at enhancing tax compliance. This experimental study is a replication and an extension of the study of Fochmann et al. (2018), who find that fully prefilled forms have a significant influence on compliance behavior. In this study, I focus on the situation of partial prefilling and investigate the influence of partly prefilled tax returns on compliance behavior in a controlled experiment. I find that partial prefilling with correct values increases compliance for the prefilled fields compared to the non-prefilled fields. In case of incorrect prefilling, I do not observe a difference in compliance for the prefilled fields compared to the non-prefilled fields. Additionally, I show that the findings of Fochmann et al. (2018) regarding the adjustment behavior for prefilled values can be transferred to different prefilling scenarios.

Keywords

Prefilling, Dishonesty, Prefilled Forms, Tax Compliance, Behavioral Economics

JEL-Classification

C91, D14, H26

Acknowledgements: I thank Martin Fochmann, André Hoppe, Frederik Schildgen and Manuel Schwaninger for helpful comments and suggestions.

Chapter 3

Tables

Figures

3.1 Introduction

The importance of prefilled forms in the time of digitalization is steadily increasing. As the digital transformation is step by step taking root in all areas of life, governments are trying to keep up and implement digital processes in the different areas of governmental responsibilities. Taxation is no exception to that. Many countries, such as Canada, Australia, UK, Sweden, Norway, Denmark, Finland, Germany, Italy, France, Spain, use prefilling in the tax declaration process (OECD, 2017). Digital data exchange allows tax authorities to receive data on the taxpayer from, for example, employers, banks, and social insurance agencies, which in turn can be prefilled for the taxpayer in her tax forms. With prefilling tax returns for the taxpayer, states aim at supporting the taxpayer in filing their tax returns and provide an image of service orientation and cooperation. At the same time, the data exchange allows tax authorities to have more control over the tax relevant information of a taxpayer, and aims at improving tax compliance (OECD, 2017).

Surprisingly, so far there is only little research on the effects of prefilling. Some studies provide first important insights using empirical data on third-party reporting (Kleven et al., 2011, Kotakorpi and Laamanen, 2016, Gillitzer and Skov, 2016). Fochmann et al. (2018) are first to study the influence of prefilling on compliance in a lab experiment. They examine the effects of prefilled tax returns on compliance behavior in situations where individuals receive a fully prefilled tax return (or a fully blank tax return). The fully prefilled tax return can be correctly prefilled or incorrectly prefilled, and if incorrectly prefilled, leading to either tax advantages or tax disadvantages. They find that prefilling influences tax compliance behavior of taxpayers. In particular, they observe that prefilling can enhance compliance compared to non-prefilling, but that the quality of prefilling matters. Correct prefilling enhances compliance, but incorrect prefilling does not increase compliance compared to non-prefilling, if the incorrect values lead to tax savings for the taxpayer. This study builds on the research of Fochmann et al. (2018) (henceforth FMO) and has two aims: first, replicating their findings in different prefilling scenarios, and second, providing additional analyses on the compliance behavior and the adjustment behavior for prefilled values.

The contribution of my study is as follows. First, my study scrutinizes the effects of partial prefilling of tax returns on compliance behavior. This experiment extends the original study, which focuses on a between-subject comparison, by a within-subject analysis. If tax returns are prefilled, it might often be the case that they are not completely prefilled. For example, in the case of third party reporting, information from employers, banks and social insurance agencies

are available and can be prefilled, but other types of income sources, such as income from renting out property or from self-employed activities, need to be reported by the taxpayer herself. It is important to study the effects of prefilling in the different tax filing situations that taxpayers face, and I contribute to the understanding of compliance behavior by providing insights on how the effects of prefilling found by FMO can be transferred to the situation of partial prefilling. Complementing the research on the effects of prefilling in different settings contributes to the tax compliance literature and might provide further practical insights and policy implications for governments.

Second, I provide a detailed analysis on the adjustment behavior of taxpayers for prefilled values. When taxpayers receive a tax return with prefilled values, these values can be adjusted before submitting the tax return. FMO find that if the prefilled values are incorrect, individuals are aware of that and adjust these values. The amount of adjustment is depending on the deviation of the prefilled value from the true income. In their study, incorrect prefilling is always combined with correct prefilling in one tax return. I aim to replicate their findings in different prefilling scenarios. I extend their analysis of adjustment behavior by separating incorrect and correct prefilling and combining these with non-prefilling. Furthermore, I investigate the case that different prefilling errors occur in one tax return, i.e., the tax return contains both errors leading to tax savings as well as errors leading to extra tax burdens. Thus, I contribute to the understanding of compliance behavior of taxpayers by providing further insights on the adjustment behavior of prefilled (especially incorrectly prefilled) values in different prefilling scenarios.

Additionally, my study contributes to the literature on replication in experimental economics. In recent years, various researchers have highlighted the importance of replication for research and emphasize the need for replication studies in experimental economics (e.g., Rosenblat et al., 2015, Camerer et al., 2016). Replicating results makes research more reliable and can be seen as a core element of any scientific process (e.g., Schmidt, 2009, Eckel et al., 2015). It might also be interpreted as a quality hallmark (e.g., Vallois and Jullien, 2017). Furthermore, replication tests the robustness of findings and can provide valuable insights on how these findings translate to other settings and situations.

In the literature, replication can have different meanings and names. A study can attempt to directly replicate another paper, which can be called, for example, “direct replication”, “reproduction”, or “pure replication” (see, for example, Vallois and Jullien, 2017). A pure replication study aims at generating the same (or similar) results by using the same experimental protocol. Furthermore, a study can “incorporate replication as a basis for extending the original

work in a particular direction" (Deck et al., 2015, p. 4), which can be called, for example, "replication and extension", but also be declared as a "robustness test" (e.g., Vallois and Jullien, 2017, Clemens, 2017). Replication and extension studies aim at examining if the original results can be transferred to a new context. Extensions can be the variation of parameters of the original study, changes in the environmental circumstances, or the implementation of new treatments. My study is related to this type of replication, as I study whether the findings of FMO can be transferred to different prefilling scenarios.

I conduct a lab experiment, in which individuals receive a partly prefilled tax return: half of the information is already prefilled, and the other half is not prefilled, thus needs to be filled in by the taxpayer herself. I consider the following cases: a tax return can be partly prefilled with correct information or partly prefilled with incorrect information, whereas the incorrect information can either lead to tax savings (when the prefilled income is lower than the true income, thus leading to a lower tax liability) or tax disadvantages (when the prefilled income is higher than the true income, thus leading to a higher tax liability) for the taxpayer. Additionally, I consider the case that one tax return contains different kinds of errors, thus, the tax return is prefilled with incorrectly lower and also incorrectly higher values. As in the original study, all monetary aspects are kept constant and the only variation is the prefilling in the treatments.

My results are as follows. First, I find cautious evidence that partial prefilling with correct information increases compliance in comparison to the non-prefilled fields of a tax return. This finding corroborates the findings of FMO, who observe a higher tax compliance level for correctly prefilled tax returns compared to non-prefilled tax returns. Second, in the case of partial prefilling with incorrect values, I do not observe a difference in compliance for the prefilled fields compared to the non-prefilled fields, independent of whether the error leads to tax savings or tax disadvantages for the taxpayer. For the case of lower prefilled values, this finding supports the findings of FMO, who do not observe a difference in compliance for incorrectly lower prefilling compared to non-prefilled tax returns. Nevertheless, this finding contradicts their finding of a higher compliance level for incorrectly higher prefilling than for blank tax returns.

Third, with regard to the adjustment behavior, I can replicate the findings of FMO for all my prefilling scenarios. In accordance with the original study, I find that individuals are aware of the incorrect prefilling and adjust almost all of the incorrectly prefilled values, whereas the majority of the correctly prefilled values are accepted and not adjusted. Interestingly, for the incorrectly prefilled values, the amount of adjustment is purely depending on the size and the direction of the concrete prefilling error and not on the overall prefilling scenario of the

taxpayer. In line with the original study, I observe that in the case of incorrectly lower prefilling (which would result in tax savings for the taxpayer), the prefilled value is only adjusted if it largely deviates from the true value. Individuals do not adjust the incorrectly lower value if the deviation is small. In contrast, individuals (almost) always adjust the prefilled values if the prefilled value is incorrectly high and would impose a tax disadvantage on the taxpayer. These findings are independent of my different prefilling scenarios.

The results of my study might have practical implications. For example, tax administrations might not always be able to provide prefilled forms with full information for all taxpayers. Depending on different types of income sources, only partial prefilling might be feasible. Even if further information is required, partial prefilling can induce a positive effect on compliance, but my results suggest that this positive effect is not to be expected if prefilling is associated with errors. However, in this case, compliance is still on the same level as for non-prefilling. Consequently, prefilling is preferable over non-prefilling also in the case that only partial information can be provided. Furthermore, my results suggest that the observed adjustment behavior for prefilling errors are robust behavioral patterns that can be transferred to different prefilling scenarios, which might offer practical insight and policy implications for tax administrations.

The remainder of this paper is structured as follows. Section 2 describes the experimental design. Section 3 presents the hypotheses. The compliance behavior is analyzed in section 4, the adjustment behavior in section 5. Section 6 concludes.

3.2 Experimental Design

3.2.1 Experimental Design and Treatments

For this study, I use exactly the same experimental design as in the original study.[24] The instructions are provided to the participants at the beginning of the experiment (see appendix A1). In the first part of the experiment, participants have the possibility to earn their pre-tax income in a real effort task in six separate rounds. In the second part of the part of the experiment, participants must declare their income in a tax return that consists of six income fields, one for each round of the real effort task. The only difference in the design is the treatment variation.

Differently to FMO, I consider the case that participants receive a partly prefilled tax return. In the partly prefilled tax return, three of the six income fields are prefilled and the other

[24] See Fochmann et al. (2018) for a detailed description of the experimental design.

three income fields need to be filled-in by the taxpayer herself. This design allows me to analyze the differences in compliance between prefilling and non-prefilling in a within-subject comparison. Contrary to the original study, which combines incorrect prefilling always with correct prefilling in one tax return, I separate correct and incorrect prefilling in all my treatments. In the partly prefilled tax returns, the prefilled values are either correct (treatment B+CP) or incorrect. If the prefilling is incorrect, the prefilled values are either lower (treatment B+LP) or higher (treatment B+HP) than the true income, leading to either tax savings or tax disadvantages, respectively. Only in my last treatment, participants receive a fully incorrectly prefilled tax return, which contains both incorrectly lower and incorrectly higher prefilled values. This design allows me to analyze the effects of different kinds of errors in one tax return. Figure 1 presents the experimental design and the treatment differences.

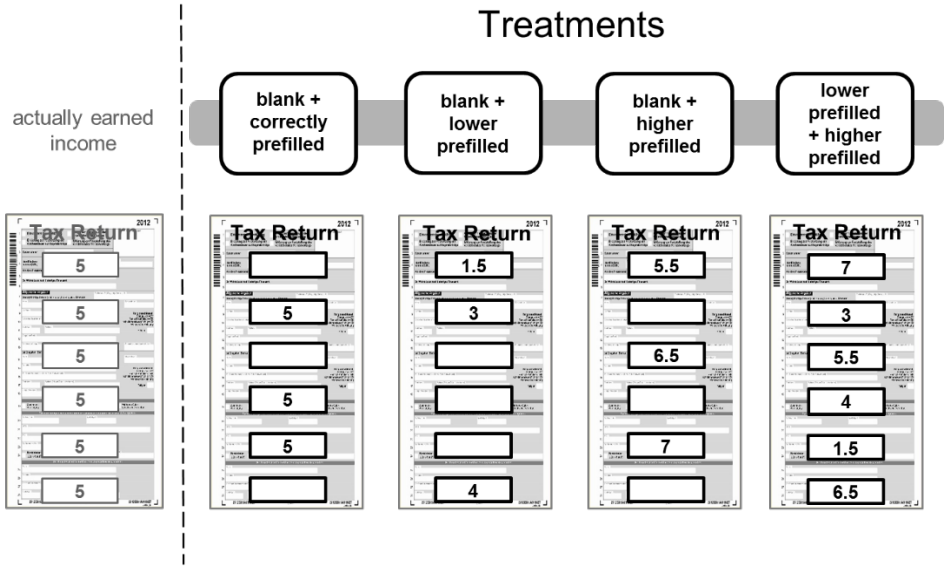


Figure 1: Experimental Design

Note: This figure highlights the differences among the treatments. Subjects are randomly distributed to one of the four treatments.

- *Treatment B+CP (**b**lank + **c**orrectly **p**refilled)*: 3 income fields of the tax return are not prefilled (*blank);* the other 3 income fields of the tax return are *correctly* prefilled (i.e., prefilled income equals true income).
- *Treatment B+LP (**b**lank + **l**ower **p**refilled)*: 3 income fields of the tax return are not prefilled (*blank);* the other 3 income fields of the tax return are *incorrectly lower* prefilled to the advantage of the taxpayer (i.e., prefilled income is *lower* than true income).

- *Treatment B+HP (**b**lank + **h**igher **p**refilled)*: 3 income fields of the tax return are not prefilled (*blank);* the other 3 income fields of the tax return are *incorrectly higher* prefilled to the disadvantage of the taxpayer (i.e., prefilled income is *higher* than true income).
- *Treatment LP+HP (**l**ower **p**refilled + **h**igher **p**refilled)*: 3 income fields of the tax return are *incorrectly lower* prefilled to the advantage of the taxpayer (i.e., prefilled income is *lower* than true income); the other 3 income fields of the tax return are *incorrectly higher* prefilled to the disadvantage of the taxpayer (i.e., prefilled income is *higher* than true income).

For the partly prefilled tax returns, I randomly vary which three of the six income fields are prefilled. For the last treatment, I also randomly vary which three income fields are lower (higher) prefilled. This design allows to analyze if there are order effects influencing compliance behavior in the six income fields. Furthermore, if the prefilled income is incorrect, the size of error, i.e., the deviation of the incorrect values from the correct values, is randomly varied and equally distributed between 0% and 100% of the true income. Thus, the lower prefilled values lie between 0% and 100% of the true income, and the higher prefilled values lie between 100% and 200% of the true income. This variation allows to analyze how the size of error influences individual tax compliance behavior.

All other factors that may influence tax compliance, such as audit probability, fine, and tax rate, are kept constant between the different treatments. The only variation between the treatments is the prefilling of the tax returns. As in the original study, if there are differences in compliance, they can only result from the prefilling variations.

3.2.2 Sample and Data

The experiment was conducted at the computerized experimental laboratory of the University of Cologne (CLER) from June to October 2018. Participants were recruited with ORSEE (Greiner, 2004) and the experimental software I used was z-Tree (Fischbacher, 2007). In total, 158 subjects (mainly undergraduate students, 88 females and 70 males) participated and earned, on average, EUR 20.42 in approximately 96 minutes (approximately EUR 12.76 per hour). A total of 40 subjects were randomly assigned to treatments B+CP, B+LP and B+HP each and 38 to treatment LP+HP. Over all treatments, I have 948 observations (158 subjects with 6 decisions per subject). Table 1 provides an overview of the participants' main characteristics.

Table 1: Main Characteristics of Participants

Variable	description	mean
female	female = 1; male = 0	55.70%
risk attitude	Holt&Laury (2002) risk measure	3.99 / 10
age	in years (18 to 68)	25
economics	study with more than one lecture in economics = 1 (0 otherwise)	45.57%
bachelor	education with a bachelor's degree = 1 (0 otherwise)	55.70%
tax experience	experience with tax returns = 1 (0 otherwise)	43.04%
tax knowledge	tax knowledge = 1 (0 otherwise)	15.19%
income	in Euro (monthly income after fixed costs)	370.10
tax morale	0 to 9; low tax morale = 0; high tax morale = 9	6.85
fairness	0 to 10; low perceived fairness of tax and control system in experiment = 0; high perceived fairness of tax and control system in experiment = 10	6.57
decision complexity	0 to 10; low perceived decision complexity in experiment = 0; high perceived decision complexity in experiment = 10	1.91
joy	0 to 10; felt no joy during experiment = 0; felt high joy during experiment = 10	4.70
anger	0 to 10; felt no anger during experiment = 0; felt high anger during experiment = 10	4.24
fear	0 to 10; felt no fear during experiment = 0; felt high fear during experiment = 10	1.87
guilt	0 to 10; felt no guilt during experiment = 0; felt high guilt during experiment = 10	1.15

Note: This table provides an overview of the individual characteristics of the 158 participants in the experiment.

3.3 Hypotheses

Traditional economic theory models tax compliance based on monetary factors, such as the tax rate, audit probability and fine (Becker, 1968, Allingham and Sandmo, 1972). Building up on this research, in recent years, various researchers consider psychological factors and study the effects of, for example, social norms, fairness and moral costs on tax compliance behavior (see, for example, Bosco and Mittone, 1997, Fortin et al., 2007, Hofmann et al., 2008, Dulleck et al., 2016). In this experiment, the prefilling variation does not influence the monetary factors of the tax compliance decision. The tax rate, audit probability and fine are kept constant in the different prefilling variations. However, prefilling might influence compliance behavior via non-monetary factors.

The influence on compliance via non-monetary factors might work through several psychological phenomena. First, prefilling might influence the moral costs of tax evasion, and an increase or a decrease in moral costs can influence compliance behavior in different

directions (see, for example, Erard and Feinstein, 1994, Frey and Torgler, 2007, Kirchler, 2007). Second, the prefilled values might serve as anchors in the compliance decision and therefor affect compliance via the anchoring effect (Tversky and Kahneman, 1974). Furthermore, the default effect might influence tax compliance behavior (e.g., Mazar et al., 2015), as the prefilled values might serve as defaults for the income that needs to be declared. Again, the anchors as well as the defaults might influence compliance in different directions.

As this study aims at replicating the findings of FMO, my hypotheses build on the results of their study regarding the effects of prefilling on compliance behavior and on adjustment behavior. For a very detailed discussion of how the mentioned psychological effects can influence compliance behavior in the different cases of prefilling, see FMO.

Compliance Behavior

First, the original study finds that tax compliance is higher for correctly prefilled tax returns than for blank tax returns in a between-subject comparison. They argue that correct prefilling increases the moral costs of tax evasion (as a deliberate adjustment of correct values is necessary), which results in a higher tax compliance level. Furthermore, correctly prefilled values might increase tax compliance via the anchoring effect. I expect to observe the same result for the situation that taxpayers receive tax returns which are only partly prefilled with correct values and formulate the following hypothesis:

Hypothesis 1: *The tax compliance level is higher for the correctly prefilled income fields than for the blank income fields in a partly prefilled tax return.*

Second, I examine the effects of incorrectly lower prefilled values, which lead to tax savings for the taxpayer, on compliance compared to non-prefilled income fields. FMO compare the tax compliance level for incorrectly lower prefilled values with correct prefilling. They argue that the anchoring effect might reduce compliance for the lower prefilled fields in comparison to correctly prefilled fields and that incorrectly lower prefilling reduces the moral costs of tax evasion (license to cheat, responsibility shifting, omission bias), which also might reduce the tax compliance level. Comparing incorrectly lower prefilled income fields with blank tax returns, they find no difference in the tax compliance level (between-subject comparison). Therefore, I expect that the tax compliance level does not differ in a within-subject setting, i.e., for the incorrectly lower prefilled income fields and the non-prefilled income fields in one tax return.

Hypothesis 2: *The tax compliance level does not differ between the incorrectly lower prefilled income fields and the blank income fields in a partly prefilled tax return.*

Third, I consider the case that incorrectly higher prefilled values, which lead to tax disadvantages for the taxpayer, are combined with non-prefilled income fields in one tax return. FMO argue that the tax compliance level might be higher due to the anchoring effect, but that the moral costs of tax evasion might be lower due to incorrect prefilling (license to cheat), which would reduce tax compliance. Also a loss of trust or perceived unfairness might reduce compliance (Hofmann et al. 2008, Kirchler et al. 2008). They conclude that the total effect is unclear. They find a higher tax compliance level for the incorrectly higher prefilled income fields (in combination with correct prefilling in one tax return) compared to blank tax returns. I expect to replicate this finding in a within-subject design, i.e., in a situation where the taxpayer receives a tax return that is only partially prefilled with incorrect values leading to tax disadvantages, and formulate the following hypothesis:

Hypothesis 3: *The tax compliance level is higher for the incorrectly higher prefilled income fields than for the blank income fields in a partly prefilled tax return.*

Finally, I extend the analysis of compliance behavior for incorrect prefilling and study the effects of different kinds of errors in one tax filing situation. The tax return contains both incorrect values that lead to tax savings, and incorrect values that lead to tax disadvantages for the taxpayer. FMO find that the tax compliance level for incorrectly higher prefilled income fields is higher than for incorrectly lower prefilled income fields in a between-subject comparison (in both cases, incorrect prefilling is combined with correct prefilling). Based on this result, I expect the tax compliance level to be higher for the incorrectly higher prefilled values than for the incorrectly lower prefilled values in one tax return.

Hypothesis 4: *The tax compliance level is lower for the incorrectly lower prefilled income fields compared to the incorrectly higher prefilled income fields in one tax return.*

Adjustment Behavior

FMO find that participants actively adjust most of the incorrectly prefilled values, whereas the majority of the correctly prefilled values are not adjusted and declared as prefilled. I expect to find the same behavioral pattern in my treatments. Thus, I expect that if a tax return is partly prefilled with correct information, most of these values are not adjusted, whereas for the partly incorrectly prefilled tax returns, taxpayers actively adjust most of the prefilled values. For the combination of higher and lower prefilling, I consequently expect that almost all income fields of the tax return are adjusted. I formulate the following hypotheses:

Hypothesis 5: *Subjects adjust the majority of incorrectly prefilled income fields, whereas they do not adjust the majority of the correctly prefilled values.*

Furthermore, the original study finds that almost all incorrectly higher prefilled values are adjusted downward, whereas for the lower prefilled income fields, the majority is adjusted upward. In the case of higher prefilled values, the downward adjustment overcompensates the prefilling error, but for lower prefilled values, the upward adjustment does not compensate the error. I expect to replicate this finding for the partly prefilled tax returns, and also for the case that a tax return is prefilled with incorrectly high and incorrectly low values.

Hypothesis 6: *The majority of higher prefilled values are adjusted downward, whereas the majority of lower prefilled values are adjusted upward. The downward adjustment of higher prefilled values overcompensates the incorrect prefilling, whereas the upward adjustment of lower prefilled values does not compensate the prefilling error.*

Moreover, the original study finds that the adjustments made by subjects depend on the level of deviation of the incorrect values from the true value (i.e., size of error). They find that subjects do not adjust lower prefilled values, if the deviation is rather small (deviations lower or equal to 50% of the true value), but that they adjust them upward for large deviations (deviations higher than 50% of the true value). In contrast to that, they observe that higher prefilled values are (almost) always adjusted downward.

Hypothesis 7: *For small deviations, subjects do not adjust lower prefilled values. Only for large deviations, subjects adjust lower prefilled values upward. Higher prefilled values are almost always adjusted downward.*

3.4 Results: Compliance Behavior

3.4.1 Declared Share of True Income

The first measure analyzed is the declared share of true income. This relative measure is calculated as the declared income in relation to the true income for each single income field and thus, accounts for income heterogeneity. If the taxpayer declares the true income, she acts tax compliant. If she declares an income that is lower (higher) than the true income, she acts in a tax evasive (over-compliant) way.

Descriptive Statistics and Nonparametric Statistics. As the taxpayer declares income in six income fields in the tax return, there are six observations per individual. For each individual, an average tax compliance level is calculated. This was done separately for the non-prefilled and for the prefilled (in treatment LP+HP separately for the lower and the higher prefilled) income fields. Figure 2 shows the mean values of declared share of income for each treatment.[25]

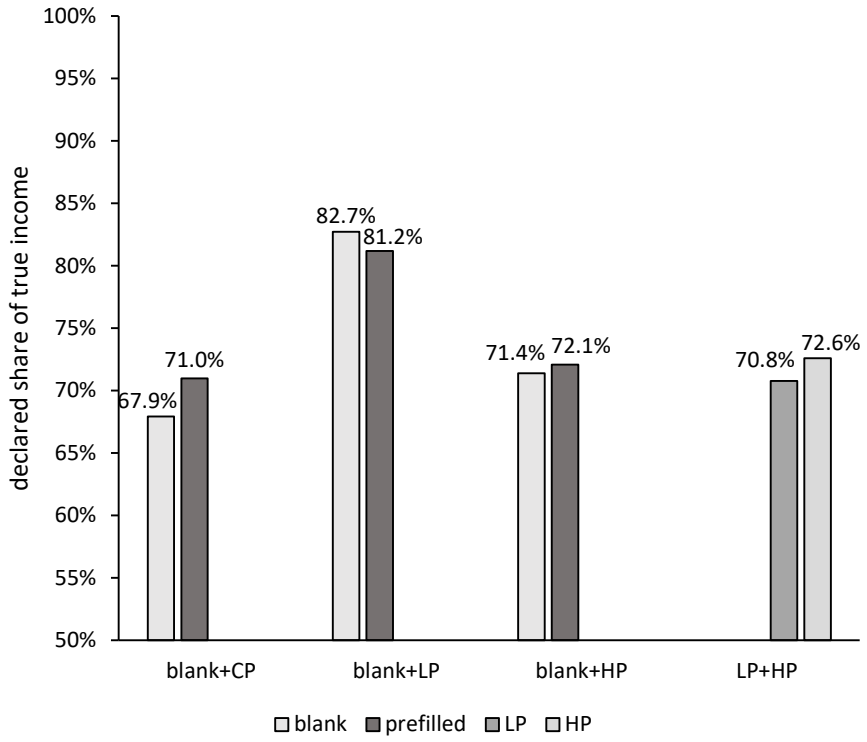


Figure 2: Declared Share of Income in the Treatments

Note: This figure shows the relative declared income for each treatment on average separately for blank and prefilled income fields (separately for lower and higher prefilled income fields in treatment LP+HP). For each individual, an average tax compliance level is calculated.

[25] In line with FMO, I exclude observations with a declared income greater than 200% of the actually earned income. For the incorrectly prefilled values, the largest deviation possible is 100% of the true income. Thus, incorrectly prefilled values can range from 0% to 200% of the true income. There is no reason for a declared income higher than 200% of the actually earned income. All stated results will stay the same if I include all observations in the analyses.

The mean declared share of true income lies between 67.9% and 82.7% in the four treatments. In line with the experimental tax compliance literature, I find an average tax compliance level between 0% and 100%.

In treatments B+CP, B+LP and B+HP I find no significant differences in a within-subject comparison for the mean declared share of true income for the blank income fields compared to the prefilled income fields (Wilcoxon signed-ranks, all p-values > 0.1, two-tailed). In treatment LP+HP, I also find no significant difference for the mean declared share of true income for the lower prefilled income fields compared to the higher prefilled income fields (Wilcoxon signed-ranks, p-value > 0.1, two-tailed).

Regression Analyses. In addition to the nonparametric statistics, I run linear regressions for the within-subject comparisons. The dependent variable is the declared share of true income. For the regression analyses, each observation of the six income fields of each subject is considered. I use subject-specific random effects to account for dependency of the six decisions on subject-level. Furthermore, subject characteristics such as gender, risk attitude, age, economics background etc. are included in the regressions.[26] The regression results are presented in Table 2. For models 1–3, the variable of interest is the prefilled variation. The dummy variable "prefilled" takes the value of one if the declared share of income refers to a prefilled income field (0 otherwise). The coefficients of the prefilled variable measure the difference in the declared share of true income between the blank income fields and the prefilled income fields. Thus, models 1–3 study whether prefilling has an effect on the tax compliance level in comparison to non-prefilling. In model 4, the variable of interest is the lower prefilled variation. The dummy variable "lower prefilled" takes the value of one if the declared share of income refers to a lower prefilled income field (0 otherwise). The coefficient of the lower prefilled variable measures the difference in the declared share of true income of the lower prefilled compared to the higher prefilled income fields in treatment LP+HP.

In the regression analysis for treatment B+CP, I find a significant positive coefficient for the prefilled variable (p-value < 0.1). Thus, I observe a significant higher tax compliance level for the correctly prefilled income fields compared to the non-prefilled income fields. As the higher tax compliance level for the correctly prefilled income fields is not statistically significant in the non-parametric test, the regression results should be seen as a cautious

[26] In all models, I included the 15 individual specific variables reported in table 1 in the regressions. The individual variables are not displayed, but the complete set of all regression results is presented in table A2.1 in the appendix.

evidence for a positive effect of correct prefilling on the tax compliance level in a within-subject comparison.

***Result 1**: Hypothesis 1 is supported. I find cautious evidence for a higher tax compliance level for the correctly prefilled income fields compared to the non-prefilled income fields in the case of partly prefilled tax returns.*

In the regression analysis for treatment B+LP, I do not observe a significant coefficient for the prefilled variable. Thus, as in the original study, I do not observe a difference in the tax compliance level between the incorrectly lower prefilled and the blank income fields (treatment B+LP) in the regression analysis, which supports the findings of the non-parametric statistics.

***Result 2:** Hypothesis 2 is supported. I find no difference in the compliance level for incorrectly lower prefilled income fields compared to non-prefilled income fields in the case of partly prefilled tax returns.*

In contrast to the findings of the original study, I do not find a significant difference in the tax compliance level between the incorrectly higher prefilled and the blank fields (treatment B+HP) in the regression analysis, which supports my findings in the non-parametric statistics.

***Result 3:** Hypothesis 3 cannot be supported. I do not observe a difference in the compliance level for incorrectly higher prefilled income fields compared to non-prefilled income fields in the case of partly prefilled tax returns.*

Also, I do not observe a significant coefficient for the lower prefilled variable in treatment LP+HP. As in the non-parametric statistics, I do not find a significant difference in the tax compliance level between the incorrectly lower and the incorrectly higher prefilled income fields in one tax return.

***Result 4:** Hypothesis 4 cannot be supported. I find no difference in the compliance level for incorrectly lower prefilled income fields compared to the incorrectly higher prefilled income fields in one tax return.*

Table 2: Random Effects Linear Regressions, Within-Subject Comparison (dependent variable: declared share of true income)

	model 1	model 2	model 3	model 4
	blank + correctly prefilled	blank + lower prefilled	blank + higher prefilled	lower prefilled + higher prefilled
prefilled	0.03*	-0.02	0.00	
	(0.02)	(0.02)	(0.02)	
lower prefilled				-0.02
				(0.02)
individual controls	yes	yes	yes	yes
constant	0.51	0.98	0.26	0.33
	(0.55)	(0.87)	(0.74)	(1.06)
no. of observations	235	233	235	220
no. of subjects	40	39	40	37
R-squared				
within	0.0186	0.0052	0.0000	0.0032
between	0.3878	0.3985	0.6197	0.4842
overall	0.3658	0.3641	0.5216	0.4159

Note: In this table, the results of random effects linear regressions are presented with the declared share of true income as dependent variable (regression coefficients, standard errors in parentheses). To analyze the differences within the treatments in models 1-3, I use the dummy variable "prefilled", which takes the value 1 if the declared income refers to a prefilled income field (0 otherwise). To analyze the differences within the treatment in model 4, I use the dummy variable "lower prefilled", which takes the value 1 if the declared income refers to a lower prefilled income field (0 otherwise). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

3.4.2 Share of Compliant Decisions

A second measure that is analyzed is the share of compliant decisions. For each income field, the taxpayer has to decide, if she declares the true income, thus, if she acts in a compliant way, or if she declares an income not matching the true income and consequently not being tax compliant. As there are six income fields in the tax return, a taxpayer makes the compliance decision, i.e., the decision to act compliant or not, six times. For each individual, an average share of compliant decision is calculated, thus, the share of compliant decisions shows, for how many of her six decisions the taxpayer chooses to be compliant (relatively). This relative measure is calculated separately for the non-prefilled and for the prefilled (in treatment LP+HP separately for the lower and the higher prefilled) income fields. Figure 3 shows the share of compliant decisions on average for each treatment.

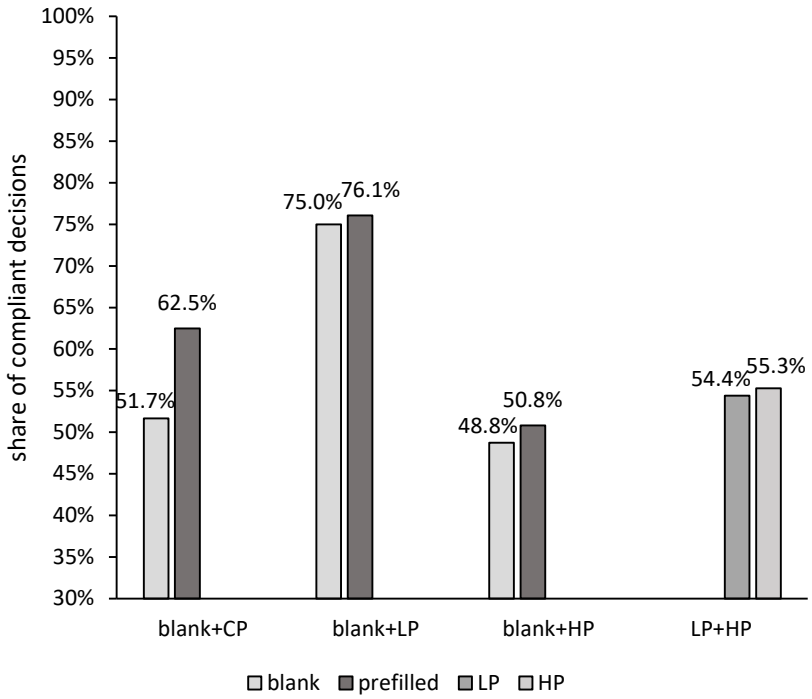


Figure 3: Share of Compliant Decisions in the Treatments

Note: This figure shows the share of compliant decisions for each treatment on average separately for blank and prefilled income fields (separately for lower and higher prefilled income fields in treatment LP+HP). An average share of compliant decisions for each subject is used.

Descriptive Statistics and Nonparametric Statistics. The mean share of compliant decisions lies between 48.8% and 76.1% in the four treatments. In treatment B+CP, the share of compliant decisions for the correctly prefilled income fields is 62.5% and is significantly higher than for the blank income fields (51.7%) (Wilcoxon signed-ranks, p-value < 0.1, two-tailed). Thus, if the income fields are correctly prefilled, a taxpayer chooses to be compliant significantly more often than for non-prefilled income fields.

For all other treatments (B+LP, B+HP, LP+HP) I find no significant differences in a within-subject comparison for the mean share of compliant decisions (Wilcoxon signed-ranks, all p-values > 0.1, two-tailed). Thus, I find no differences regarding an individual's decision to be compliant or not between non-prefilled income fields and incorrectly prefilled income fields, independent of the type of error.

Regression Analyses. In addition to the nonparametric statistics, I run logit regressions for the within-subject comparisons. For the regression analyses, each of the six decisions of each subject is considered. The dependent variable now is the dummy variable compliance decision, which takes the value of one, if the taxpayer chooses to act compliant, i.e. declares her true income, in that income field (0 otherwise). I use subject-specific random effects to account for dependency of the six decisions on subject-level. In addition, subject characteristics such as

gender, risk attitude, age, economics background etc. are included in the regressions.[27] The regression results are presented in Table 3. For models 1–3, the variable of interest is the prefilled variation. The dummy variable "prefilled" takes the value of one if the compliance decision refers to a prefilled income field (0 otherwise). Thus, models 1–3 study whether prefilling has an effect on the decision to act compliant or not. In model 4, the variable of interest is the lower prefilled variation. The dummy variable "lower prefilled" takes the value of one if the compliance decision refers to a lower prefilled income field (0 otherwise).

In the regression analysis for treatment B+CP, I find a significant positive effect of the prefilled variable (p-value < 0.01). Thus, in line with the non-parametric statistics, I find that taxpayers significantly more often choose to declare their true income, if the income is correctly prefilled. Consequently, correct prefilling enhances compliance. This finding supports the findings summarized in result 1. For all other treatments (B+LP, B+HP, LP+HP) I find no significant differences in a within-subject comparison for the compliance decision (Wilcoxon signed-ranks, all p-values > 0.1, two-tailed), which supports the findings of the non-parametric statistics. These findings are in line with my findings regarding the average tax compliance level and support the findings summarized in results 2-4.

Table 3: Random Effects Logit Regressions, Within-Subject Comparison (dependent variable: compliance decision)

	model 1	model 2	model 3	model 4
	blank + correctly prefilled	blank + lower prefilled	blank + higher prefilled	lower prefilled + higher prefilled
prefilled	2.29***	-0.28	0.50	
	(0.73)	(0.75)	(0.53)	
lower prefilled				-0.08
				(0.52)
individual controls	yes	yes	yes	yes
constant	-4.40	78.45*	-0.79	13.43
	(8.76)	(45.57)	(11.49)	(13.30)
no. of observations	235	233	235	220
no. of subjects	40	39	40	37

Note: This table presents the results of random effects logit regressions with "compliance decision" as dependent variable (standard errors in parentheses). The dummy variable "compliance decision" takes the value 1 if the taxpayer acts compliant, i.e., declares her true income (0 otherwise). To analyze the differences within the treatments in models 1-3, the dummy variable "prefilled" is used, which takes the value 1 if the compliance decision refers to a prefilled income field (0 otherwise). To analyze the differences within the treatment in model 4, the dummy variable "lower prefilled" is used, which takes the value 1 if the compliance decision refers to a lower prefilled income field (0 otherwise). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

[27] In all models, I included the 15 individual specific variables reported in table 1 in the regressions. The individual variables are not displayed, but the complete set of all regression results is presented in table A2.2 in the appendix.

3.4.3 Robustness Tests

Order Effects

For the partly prefilled tax returns, I randomly varied which three of the six income fields are prefilled. Also for treatment LP+HP, I randomly varied which of the three income fields are incorrectly lower, and which are incorrectly higher prefilled. This design allows to analyze whether the compliance behavior for the different income fields is influenced by order effects. Each treatment is tested separately as well as the prefilled and the non-prefilled income fields. The same holds for treatment LP+HP for the lower and the higher prefilled income fields. I checked whether there are differences in the declared share of true income among the six income fields. I do not observe any significant differences in the compliance level for the six different income fields (Kruskal-Wallis test, all p-values > 0.1, two-tailed). Furthermore, I checked whether there are differences in the decision to act compliant or not among the six income fields, and I also do not observe any significant differences for the compliance decision among the six income fields (chi-squared test, p-value > 0.1, two-tailed). Thus, it can be excluded that the compliance behavior of subjects is influenced by order effects.

Perceived Audit Probability, Tax Morale, Fairness

In the experiment, all monetary aspects are kept constant and the prefilling variation has no influence on the actual audit probability. However, if information is already prefilled by the tax administration, individuals might have the feeling that adjusting these values increases the probability of an audit. Therefore, I examine whether the perceived audit probability differs between the treatments. In line with the procedure in FMO, I first checked whether participants were aware of the audit probability of 30% before (in the comprehension test) and after the experiment (in the ex-post questionnaire). There are only two persons out of 158 participants who gave an incorrect answer on the ex-post questionnaire. Second, the ex-post questionnaire contained the question: “How did you perceive the audit probability in the experiment?” (10-point Likert scale from “very low” to “very high”). The mean answer over the treatments is 3.5 and there are no statistically significant differences across treatments (Kruskal-Wallis test, p-value > 0.1, two-tailed). This indicates that the differences in compliance are not driven by differences in the perceived probability of an audit.[28]

[28] I also checked, if there are differences between the treatments in the number of audits actually conducted, as this might also influence the perceived probability of an audit. I do not find a significant difference across the treatments with regard to the number of conducted audits (chi-squared test, p-value > 0.1, two-tailed).

Second, I examine if there are differences regarding the tax morale of the subjects between the treatments. As in the original study, I use an adapted question from the World Values Survey in the ex-post questionnaire that is widely applied in the tax compliance literature (e.g., Slemrod 2003, Alm and Torgler, 2006): "How do you evaluate the following statement?: Cheating on tax if you have the chance..." Answers were given on a 10-point Likert scale from "…is always justifiable" = 0 to "…is never justifiable" = 9. The observed mean tax morale level in the sample is 6.9, and there are no statistically significant differences between the treatments (Kruskal-Wallis test, p-value > 0.1, two-tailed), implying differences in compliance behavior are not driven by differences in the tax morale of the subjects across treatments.

Third, I analyze if there are differences across treatments regarding the perceived fairness of the tax system in the experiment. For example, if the tax return contains errors that impose a tax disadvantage on the taxpayer, subjects might perceive the tax system as less fair. In the ex-post questionnaire, I asked: "How fair did you perceive the tax and control system that was applied in the experiment?" (11-point Likert scale from "very unfair" to "very fair"). The mean answer is 6.6 over all treatments and I observe no statistically significant differences across the treatments (Kruskal-Wallis test, p-value > 0.1, two-tailed). This indicates that differences in compliance are not driven by differences in the perceived fairness of the subjects across treatments.

Individual Variables across Treatments

For the 15 individual variables reported in table 2, I examine whether the distribution of individual characteristics differs across treatments. I do not observe that the treatments differ significantly for to most of the individual variables (chi-squared test/ Kruskal-Wallis test, p-values above 0.1, two-tailed). For example, I find no differences regarding gender or risk attitude across the treatments. Nevertheless, there are some differences regarding age (Kruskal-Wallis test, p-value < 0.01, two-tailed) and economics background (chi-squared test, p-value < 0.1, two-tailed). Pairwise comparison shows that participants in treatment B+LP are with 26.0 years (mean value) significantly older than participants in each of the other three treatments (mean values are 25.6 years in B+CP, 23.8 years in B+HP, and 23.6 in LP+HP) (Mann-Whitney U-test, p-values < 0.05, two-tailed). Regarding the economics background, in treatment LP+HP there are significantly more people with an economics background (63%) than in treatments B+LP (35%) and B+HP (40%) (Mann-Whitney U-test, p-values < 0.05, two-tailed; difference to treatment B+CP (45%) not significant, p-value > 0.1).

Consistent Tax Compliance Behavior

Furthermore, I study the consistent tax compliance behavior in the treatments. In line with the experimental tax compliance literature, in all treatments there are individuals who consistently act either fully compliant or fully evasive. I observe that 46% of all subjects consistently act compliant, thus, declare their true income in all six income fields, and 17% of the participants chose always full evasion, thus, always declared zero income. Overall, 62% of the subjects show a consistent tax compliance behavior over the six compliance decisions.[29] The distribution of subjects choosing always full evasion does not differ significantly across treatments (chi-squared test, p-value > 0.1, two-tailed), but there are differences in the distribution of subjects choosing always full compliance (chi-squared test, p-value < 0.01, two-tailed). Pairwise comparison shows that full compliance is significantly more often chosen by subjects in treatment B+LP (68%) compared to all other three treatments (45% in B+CP, 35% in B+HP, and 34% in LP+HP) (Mann-Whitney U-test, all p-values < 0.05, two-tailed).

3.5 Results: Adjustment Behavior

In this section, the adjustment behavior of taxpayers for prefilled values is analyzed in detail. I aim to replicate the findings of the original study in different prefilling scenarios and investigate whether taxpayers adjust the prefilled values, and if so, to what extent. Furthermore, I examine the adjustment behavior for incorrectly prefilled values depending on the size of the error, i.e., depending on the deviation of the incorrect value from the true value. Adjustments are defined as the difference between the declared and the prefilled income, normalized to the true income. This relative measure accounts for income heterogeneity and allows a clear comparison of the adjustment behavior between the different treatments and also of the findings in this study with the findings in the original study.

Adjustments

Table 4 shows the distribution and mean values of the adjustments of the prefilled income fields by the participants for the four treatments (for treatment LP+HP separately for the lower and higher prefilled income fields). I observe that 63% of the correctly prefilled values are not adjusted by the taxpayers. Thus, for 63% of the correctly prefilled income fields, taxpayers act tax compliant. This finding is in line with the findings of the original study, which observes that 64-70% of all correctly prefilled values are not adjusted. In contrast to that, I find that almost all incorrectly prefilled income fields are adjusted by the participants. Only 0-8% of the

[29] See table A2.3 in the appendix for treatment details.

incorrectly prefilled values in the tax returns are not adjusted. These findings are in line with the findings of the original study, where 6-7% of the incorrectly prefilled values are not adjusted. Thus, I find that taxpayers are aware of the incorrectly prefilled values and actively adjust them.

Result 5: *Hypothesis 5 can be supported. Subjects adjust the majority of incorrectly prefilled income fields, whereas they accept and not adjust the majority of the correctly prefilled values.*

Table 4: Difference between Declared and Prefilled Income (Adjustments)

adjustment		B+CP	B+LP	B+HP	LP+HP lower	LP+HP higher
downward	%	35.9%	19.0%	97.4%	28.4%	94.6%
	mean	-0.83	-0.46	-0.84	-0.53	-0.82
no adjustment	%	**63.3%**	**1.7%**	**0.0%**	**8.3%**	**2.7%**
	mean	0	0	0	0	0
upward	%	0.9%	79.3%	2.6%	63.3%	2.7%
	mean	0.20	0.47	0.31	0.53	0.02
total	%	100%	100%	100%	100%	100%
	mean	-0.30	0.28	-0.81	0.18	-0.77

Note: This table shows the distribution and mean values of the adjustments of prefilled values for the treatments (for treatment LP+HP separately for the lower and the higher prefilled income fields). Adjustments are defined as the difference between declared and prefilled income (normalized).

Direction of Adjustments

For the incorrectly lower prefilled income fields, I observe that 63–79%, thus, the majority of the lower prefilled values, are adjusted upward, whereas 19–28% are adjusted downward. The mean total adjustments amount to +0.28 in treatment B+LP and +0.18 in treatment LP+HP, which do not differ statistically (Wilcoxon signed-ranks, p-value > 0.1, two-tailed). The mean total adjustments of the prefilled values do not compensate the mean prefilling error of the incorrectly lower prefilling, which amounts to -0.47 (-0.48). In contrast, 95–97% - thus, almost all - of the incorrectly higher prefilled income fields are adjusted downward and only a negligible amount (3%) is adjusted upward. The mean total adjustments are -0.81 in treatment B+HP and -0.77 in treatment LP+HP, which do not differ statistically (Wilcoxon signed-ranks, p-value > 0.1, two-tailed). For the incorrectly higher prefilled values, the mean total adjustments overcompensate the mean prefilling error of +0.52 (+0.50). Figure 4 displays the

mean total adjustments for lower prefilled income fields (Panel A) and higher prefilled income fields (Panel B), as well as the mean prefilling error (difference between prefilled and earned income) and the difference between the declared and earned income for the different treatments. My findings support the findings of FMO, and I can formulate the following result:

Result 6: *Hypothesis 6 can be supported. The majority of lower prefilled values are adjusted upward, whereas almost all higher prefilled values are adjusted downward. The upward adjustment of lower prefilled values does not compensate the prefilling error, whereas the downward adjustment of higher prefilled values overcompensates the incorrect prefilling.*

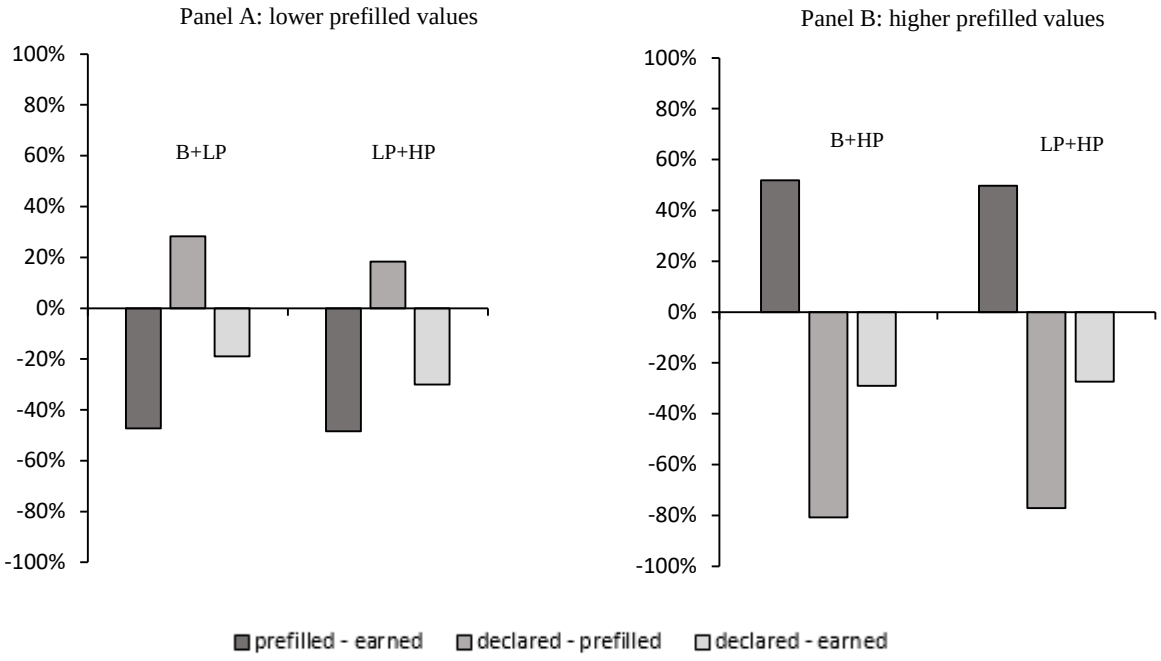


Figure 4: Adjustments of Incorrectly Prefilled Fields

Note: This figure shows the mean total adjustments (declared - prefilled) for the incorrectly prefilled income fields for lower prefilled income fields in Panel A (treatments B+LP and LP+HP) and for higher prefilled income fields in Panel B (treatments B+LP and LP+HP). Adjustments are defined as the difference between declared and prefilled income (normalized). The figure also presents the difference between prefilled and earned income and between declared and earned income (each normalized to the actually earned income).

Amount of Adjustments

Incorrectly lower prefilled values. I observe a similar adjustment behavior for the lower prefilled income fields between treatments B+LP and LP+HP. The mean downward adjustment of the lower prefilled income fields is -0.46 in treatment B+LP and -0.53 in treatment LP+HP. There is no significant statistical difference between the mean adjustments for the downward

adjustments of incorrectly lower prefilled values between the two treatments (Wilcoxon signed-ranks, p-value > 0.1, two-tailed). For the upward adjustments, the mean adjustment amounts to +0.47 in treatment B+LP and +0.53 in treatment LP+HP. Again, there is no significant statistical difference between the mean adjustments for the upward adjustments between the two treatments (Wilcoxon signed-ranks, p-value > 0.1, two-tailed). Thus, independent of the overall prefilling situation of the taxpayer, i.e., independent of the combination of incorrect values solely with non-prefilling or with incorrectly higher prefilled values, taxpayers make the same mean adjustments of incorrectly lower values. Furthermore, these findings are in line with the original study observing a mean downward (upward) adjustment for incorrectly lower prefilled values of -0.43 (+0.52) for the combination with correctly prefilled income fields.

Incorrectly higher prefilled values. I also find a similar adjustment behavior for incorrectly higher prefilled values in treatments B+HP and LP+HP. The mean downward adjustment of the higher prefilled income fields amounts to -0.84 in treatment B+LP and -0.82 in treatment LP+HP. There is no significant statistical difference between the mean downward adjustments of incorrectly higher prefilled values between treatments B+HP and LP+HP (Wilcoxon signed-ranks, p-value > 0.1, two-tailed). As only a negligible amount of upward adjustments is made (only 3 (3) of 117 (111) prefilled income fields in treatment B+HP (LP+HP)), an analysis of the amount of upward adjustments is not possible.

Thus, also for the incorrectly higher prefilled values, I observe the same adjustment behavior independent of the concrete prefilling situation of the taxpayer: independent of the combination with non-prefilling or with incorrectly lower prefilled income fields, taxpayers make the same mean adjustments of the incorrectly higher prefilled income fields. These findings are also in line with the findings in the original study. They find a mean downward adjustment of incorrectly lower prefilled values of -0.66 and a negligible amount of upward adjustments (1% upward adjustments).

Result 7: *The mean adjustments of the incorrectly lower prefilled and the incorrectly higher prefilled values do not differ between the different prefilling scenarios.*

Adjustments and Deviation

In this section, the adjustment behavior depending on the size of the error, i.e., the level of deviation of the incorrectly prefilled values from the true income, is analyzed. In a first step, I study if there are differences in the adjustment behavior for different deviation levels of the lower (higher) prefilled values between the treatments. Figure 5 shows the adjustments made by subjects depending on the deviation of the prefilled income from the true income for the different treatments.

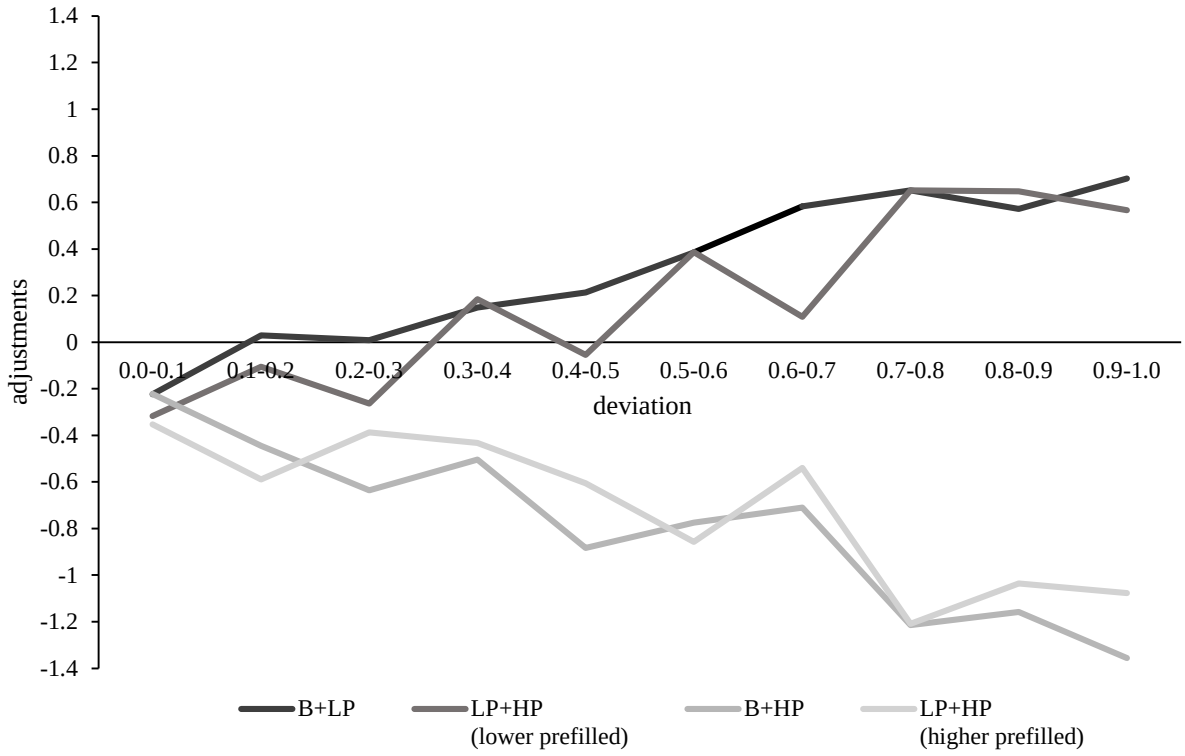


Figure 5: Adjustments and Deviation

Note: This figure shows subjects' adjustments dependent on the deviation of prefilled income from true income for the treatments. For treatment LP+HP, the adjustments are shown separately for the lower prefilled and the higher prefilled income fields. Adjustments are defined as the difference between declared and prefilled income (normalized).

Interestingly, I observe almost the same adjustment behavior for the lower (higher) prefilled income fields for the different deviation levels between treatment B+LP (B+HP) and LP+HP. For a detailed comparison, I subdivide deviation (0-100%) into ten 10%-levels, thus, I consider the mean adjustment for the deviation level of 0-10%, 10-20%, 20-30%, etc., and 90-100% in the different treatments. Comparing the mean adjustments of the lower prefilled values between treatments B+LP and LP+HP for each of the ten deviation levels, I find no statistically

significant differences (Wilcoxon signed-ranks, all p-values > 0.1, two-tailed). Thus, for each deviation level, individuals make the same mean adjustments of incorrectly lower prefilled values, independent of the prefilling scenario. Considering the higher prefilled income fields, I find almost no significant differences in the mean adjustment for each of the ten deviation levels between treatments B+HP and LP+HP (Wilcoxon signed-ranks, p-values > 0.1, two-tailed). The only exception is the mean adjustment for the deviation level of 90-100%, which amounts to -1.37 in treatment B+HP and -1.08 in treatment LP+HP and which differ significantly from each other (Wilcoxon signed-ranks, p-value < 0.5, two-tailed). Thus, except for a difference in one of the ten deviation levels for the higher prefilled values, I observe no differences in the adjustment behavior for the lower (higher) prefilled income fields for different levels of deviation between my prefilling scenarios.

Finally, I analyze the adjustment behavior for the lower (higher) prefilled income fields depending on the deviation and consider all lower (higher) prefilled values jointly (i.e., not separately for the different treatments). Figure 6 presents the adjustments made by subjects depending on the deviation for the lower prefilled income fields (treatments B+LP and LP+HP) and the higher prefilled income fields (treatments B+HP and LP+HP) and displays the mean adjustments for small deviations, i.e., deviations lower or equal to 50% of the true income, and larger deviations, i.e., deviations greater than 50% from the true value.

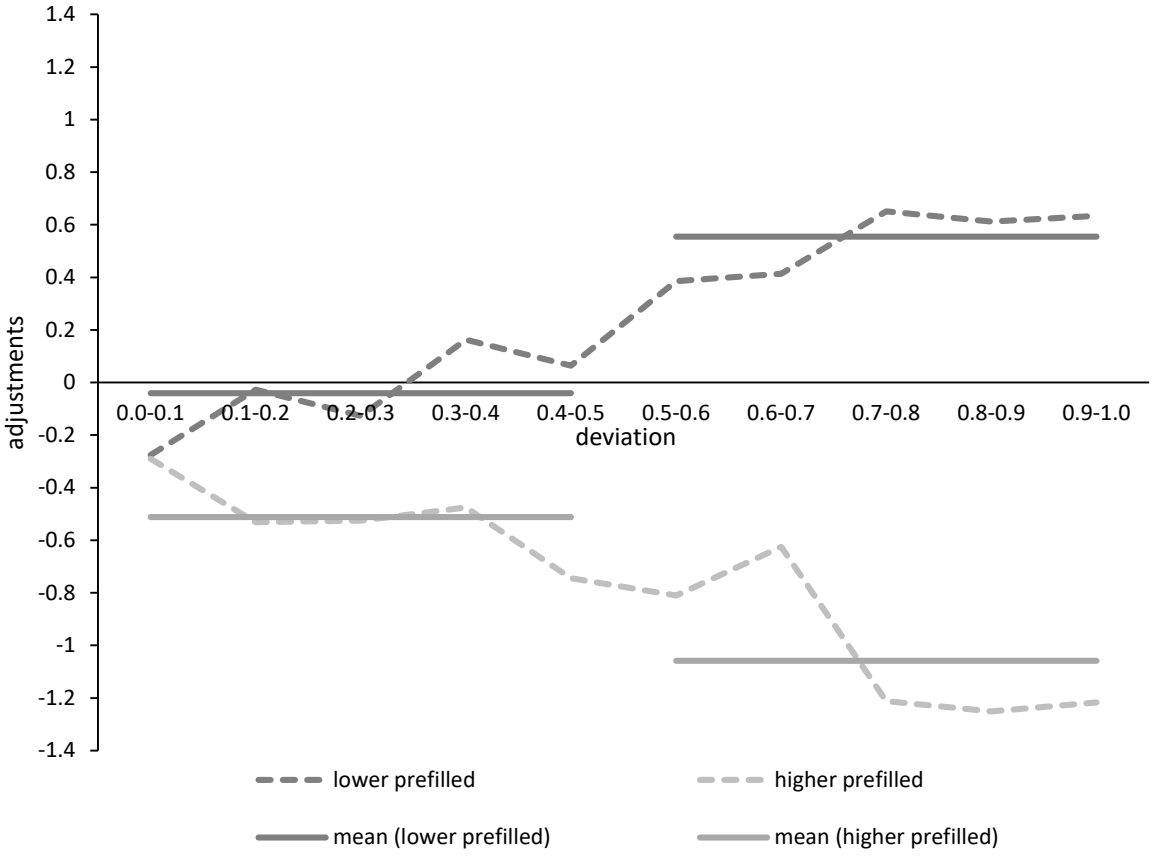


Figure 6: Mean Adjustments and Deviation

Note: This figure shows subjects' adjustments dependent on the deviation of prefilled income from true income for the lower prefilled income fields (treatments B+LP and LP+HP) and the higher prefilled income fields (treatments B+HP and LP+HP). Additionally, the mean adjustment for small deviations (0-0.5) and for large deviations (0.5-1) is displayed (again separately for the lower and the higher prefilled income fields).

Incorrectly lower prefilled values. For small deviation levels of the prefilled income (lower or equal to 50% of the true income), I observe almost no adjustments made by subjects. The mean adjustments of incorrectly lower prefilled income fields for small deviations amounts to -0.04 and does not statistically differ from zero (one sample median test, p-value > 0.1, two-tailed). In contrast to that, for larger deviation levels (greater than 50% from the true value), the mean adjustment is +0.54 and differs significantly from zero (one sample median test, p-value < 0.01, two-tailed). Thus, if the prefilling error is rather small, individuals do not adjust the incorrectly prefilled values. Only for larger prefilling errors, they adjust the incorrectly prefilled values upward. This finding is in line with the results of FMO, who also observe almost no adjustments for smaller deviations (not statistically different from zero), but a mean adjustment of +0.51 for deviations greater than 50% for the incorrectly lower prefilled income fields.

Incorrectly higher prefilled values. For incorrectly higher prefilled income, I observe that the adjustments made by individuals almost linearly relate to the size of the prefilling error: the larger the deviation of the incorrectly prefilled value from the true value, the proportionally larger the (downward) adjustment. For small deviation levels, the mean adjustment amounts to -0.51. For larger deviations, the mean adjustment is -1.02. Both adjustments differ significantly from zero (one sample median test, p-values < 0.01, two-tailed). This finding is in accordance with the results in the original study, where they also observe an almost linear relationship between the adjustments made and the deviation (they observe a mean adjustment of -0.35 (-0.80) for small (large) deviation levels, both are significantly different from zero).

Result 8: *Hypothesis 7 can be supported. Subjects make no adjustments and keep lower prefilled values if the deviation from the true value is small. For large deviations from true income, subjects adjust lower prefilled income values upward. In contrast, higher prefilled income values are almost always adjusted downward.*

Overall, I observe almost the same adjustment behavior for the lower prefilled income fields in treatments B+LP and LP+HP, and the same adjustment behavior for the higher prefilled income fields in treatments B+HP and LP+HP. These findings suggest that the observed adjustment behavior for incorrect prefilling is robust to different prefilling situations.

3.6 Discussion

Following the trend of digital transformation, tax administrations use digital data exchange in their tax declaration process and offer prefilled tax forms for the taxpayer. When tax returns are prefilled, they might be completely prefilled with all relevant information, or only parts of the form are prefilled and further information needs to be reported by the taxpayer herself. This study investigates the influence of partly prefilled tax returns on compliance behavior in a controlled experiment. My study builds on the study of Fochmann et al. (2018), who find that fully prefilled tax returns have a significant influence on compliance behavior, and aims at replicating their findings and extending their analyses in different prefilling situations.

I find that partial prefilling with correct values can increases compliance for the prefilled fields compared to the non-prefilled fields. This finding is in line with FMO, who observe a higher tax compliance level for correctly prefilled tax returns compared to non-prefilled tax

returns. Consequently, correct prefilling enhances compliance not only in the case of fully prefilled tax returns, but also in the situation that forms are only partly prefilled.

In case of incorrect prefilling, I do not observe a difference in compliance for the prefilled fields compared to the non-prefilled fields, independent of whether the error results in tax savings or tax disadvantages for the taxpayer. For the case of lower prefilled values, this finding supports the findings of the original study, who do not observe a difference in compliance for non-prefilled tax returns compared to incorrectly lower prefilling. Nevertheless, this finding contradicts their finding of a higher compliance level for incorrectly higher prefilling than for blank tax returns. An explanation for the different results might be that in the case of incorrectly higher prefilling, different non-pecuniary factors can influence compliance in different directions, and thus, the total effect on compliance is unclear. For example, according to the anchor effect (Tversky and Kahneman, 1974), higher prefilling should increase tax compliance, as the prefilled value might serve as an anchor and the taxpayer might be assimilated toward this higher value. In contrast to that, incorrect prefilling might lower the moral costs associated with tax evasion, which would lead to a lower tax compliance level (see, for example, Erard and Feinstein, 1994). Taxpayers might evaluate the adjustment of an incorrect value as less immoral than adjusting a correctly prefilled value. Moreover, errors in prefilling might be perceived by the taxpayers as a license to cheat. In addition to that, individuals might lose trust in the tax authority or feel unfairly treated by the incorrect prefilling to their disadvantage, which might further reduce tax compliance (Hofmann et al. 2008, Kirchler et al. 2008).[30] As these effects can influence compliance in different directions, the total effect on compliance is unclear, and might also be dependent on further factors. In my combination with non-prefilling or with additional types of errors, incorrectly higher prefilled values do not increase compliance. FMO combine incorrectly higher prefilling with correct prefilling in one tax return, and they find the same compliance level for the incorrectly higher and the correctly prefilled values, and thus, compared to blank tax returns, a higher tax compliance level for the incorrectly higher values.

Additionally, I show that the findings of FMO regarding the adjustment behavior for prefilled values can be transferred to different prefilling scenarios. Independent of the situation of partially incorrect prefilling or complete incorrect prefilling with different types of errors, I

[30] Furthermore, the default effect (see, for example, Mazar et al., 2008), which describes the preference of individuals to stay with a preset default option rather than actively making a decision, would suggest that higher prefilled values lead to a higher tax compliance level. However, I can exclude that the default effect influences compliance in the case of prefilling, as I find that individuals are aware of the incorrect values and adjust almost all of them.

observe the same adjustment behavior for the incorrectly prefilled values in my treatments. I find that individuals are aware of the incorrect prefilling and adjust more than 90% of the incorrectly prefilled values, whereas the majority of the correctly prefilled values are accepted and not adjusted. If the prefilled values are lower than the true values, the majority of the incorrectly prefilled values are adjusted upward, but the adjustments do not compensate the prefilling error. If the prefilled values are higher than the true income, almost all incorrectly prefilled values are adjusted downward, but in that case the adjustments overcompensate the prefilling error. Interestingly, I observe that the amount of adjustment is purely depending on the direction of the prefilling error and its size, i.e., the deviation of the incorrect value from the true value, and not on the overall prefilling scenario of the taxpayer. Lower prefilled values are only adjusted if they largely deviate from the true value. Individuals do not adjust the lower prefilled values if the error is small. In contrast to that, higher prefilled values are always adjusted downward, and the larger the error, the proportionally larger the adjustment. These results suggest, that independent of the prefilling scenario, individuals reveal the same adjustment behavior for incorrectly prefilled values.

The contribution of my study is threefold. First, my study provides insights on the effects of partial prefilling on individual compliance behavior. Tax administrations might not always have access to all tax-relevant information on a taxpayer, and in that case only partial prefilling is feasible. Therefore, it is important to study how the original findings regarding the effects of fully prefilled forms on compliance behavior can be transferred to other prefilling scenarios. Second, my study provides valuable insights on the adjustment behavior for prefilled values. Although the digital processes of data exchange and prefilling should be accurate and trustworthy, errors in these processes might occur, and these errors could lead to tax savings or tax disadvantages for the taxpayer, if they are not adjusted. I contribute to the understanding of tax compliance behavior by providing further insights on the adjustment behavior of prefilled - especially incorrectly prefilled - values in different prefilling scenarios.

Additionally, my study contributes to the literature on replication in research. Recently, considerable attention has been focused on problems of replicability of published results in all areas of science. Also in the experimental economics literature, researchers call for replication studies and highlight its importance for scientific reliability (see, for example, Rosenblat et al., 2015, Camerer et al., 2016). I contribute to the literature by providing a replication study which examines how the findings of the original study can be transferred to different prefilling scenarios.

My study uses the same experimental design as the original study, and thus, faces the same limitations resulting from this design and the usual objections regarding external validity and the subject pool of lab experiments. For a detailed discussion of the limitations, see the original study.

My study might also provide practical insights and policy implications for governments. The results suggest that correct prefilling can enhance compliance also in the case that only partial information is prefilled, but in situations where errors in prefilling occur, tax administration should not expect differences in compliance compared to non-prefilling. Nevertheless, the results suggest that prefilling is preferable over non-prefilling. Furthermore, the insights regarding the adjustments of prefilled values might be of practical relevance for tax administrations, especially in situations where errors are likely to occur, as my findings suggest that independent of the overall prefilling situation of the taxpayer, taxpayers reveal the same adjustment behavior for incorrectly lower (higher) values.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323–338.

Alm, J. and Torgler, B. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology*, 27, 224–246.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 169–217.

Bosco, L. and Mittone, L. (1997). Tax evasion and moral constraints: some experimental evidence. *Kyklos*, 50(3), 297–324.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys*, 31(1), 326–342.

Deck, C. A., Fatas E. and Rosenblat, T. (2015). Encouraging Replication of Economics Experiments. In Rosenblat, T., Fatas, E. and Deck, C. A. (eds.), Replication in Experimental Economics, 3–9.

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M. and Torgler, B. (2016). Tax compliance and psychic costs: behavioral experimental evidence using a physiological marker. *Journal of Public Economics*, 134, 9-18.

Eckel, C. C., Harwell, H. and Castillo J. G. (2015). Four Classic Public Goods Experiments: A Replication Study. In Rosenblat, T., Fatas, E. and Deck, C. A. (eds.), Replication in Experimental Economics, 13–40.

Erard, B. and Feinstein, J. (1994). The role of moral sentiments and audit perceptions in tax compliance. *Public Finance*, 49, 70–89.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.

Fochmann, M., Müller, N., Overesch, M. (2018). Less Cheating? The Effects of Prefilled Forms on Compliance Behavior. SSRN Working Paper.

Fortin, B., Lacroix, G. and Villeval, M. C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11-12), 2089–2112.

Frey, B. S. and Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of comparative economics*, 35(1), 136–159.

Gillitzer, C. and Skov, P. E. (2016). The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark. EPRU Working Paper Series.

Greiner, B. (2004). An online recruitment system for economic experiments. *Forschung und wissenschaftliches Rechnen GWDG Bericht 63*. edited by K. Kremer and V. Macho. Göttingen, Gesellschaft für Wissenschaftliche Datenverarbeitung, 79–93.

Hofmann, E., Hoelzl, E. and Kirchler, E. (2008). Preconditions of voluntary tax compliance: Knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. *Journal of Psychology*, 216, 209–217.

Kirchler, E. (2007). The economic psychology of tax behaviour. New York: Cambridge University Press.

Kirchler, E., Hoelzl, E. and Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, 29, 210–225.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.

Kotakorpi, K. and Laamanen, J. P. (2016). Prefilled income tax returns and tax compliance: Evidence from a natural experiment. Tampere Economic Working Paper 104.

Mazar, N., Amir, O. and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644.

Mazar, N. and Hawkins, S. A. (2015). Choice architecture in conflicts of interest: Defaults as physical and psychological barriers to (dis)honesty. *Journal of Experimental Social Psychology*, 59, 113–117.

OECD (2017). Third-party data management – the journey from post-assessment crosschecking to pre-filling and no-return approaches. In: Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies. OECD Publishing.

Rosenblat, T., Fatas, E. and Deck, C. A. (2015). Replication in Experimental Economics. Emerald Group Publishing.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100.

Slemrod, J. (2003). Trust in public finance. *Public Finance and Public Policy in the New Century*, ed. S. Cnossen and H.-W. Sinn, 49–88. Cambridge, USA: The MIT Press.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124–1131.

Vallois, N., Jullien, D. (2017). Replication in experimental economics: A historical and quantitative approach focused on public good game experiments. GREDEG Working Paper No. 2017-21.

Appendix

A1 Instructions

Appendix A1 includes the translated instructions (from German). All participants received the general instructions in print. Before the actual experiment was executed, subjects participated in the Holt and Laury (2002) task. The instructions for this task (first experiment) were displayed on the computer screen. After that, participants received the specific instructions for the actual (second) experiment in print.

A1.1 General Instructions

Thank you for participating in this experimental study. For your participation you receive a participation fee of 4 Euros.

The study consists of two experiments in which you have the possibility to earn money and a questionnaire at the end of the study. How much money you earn depends on your decisions and on chance. The instructions explain how you can influence how much money you earn in this study by your decisions.

It is important that you understand the instructions. Please do not hesitate to ask questions. If you have a question, please raise your hand. We will come to you to answer your question. Please do not ask your question loudly. You can write on the instructions or set markers. Please do not take the instructions home, but give them back to us at the end of the study.

The analysis of the experiment will be anonymous. We will never link your name with the data generated in the experiment. You will not learn the identity of any other participant, neither before nor after the experiment. Also the other participants will not learn your identity. At the end of the experiment, you have to sign a receipt to confirm the payments you received. This receipt will only be used for accounting purposes.

We would like to inform you that you are not allowed to communicate with other participants or leave your seat throughout the whole experiment. Please switch off your mobile phone and put it in your bag.

The calculator, the pen and the sheet of paper (for notes) that are lying on your desk, can be used.

At the end of the study you will receive your payout privately and in cash. Your total payout consists of your payout of the first experiment, plus your payout of the second experiment and the participation fee.

The instructions for the first experiment will be displayed on your computer screen.

A1.2 Instructions for the Holt and Laury (2002) Task

Please choose one of the two lotteries A or B in each of the following 10 decision situations.

You will make a decision for all 10 situations, but your payout from the first experiment is determined only by the one situation that is randomly drawn by the computer after the second experiment.

In each situation, you can either earn 2.00 € or 1.60 € from lottery A and either 3.85 € or 0.10 € from lottery B. The probabilities of winning, however, vary from situation to situation. The further down you move in the table, the higher is the probability of the higher payment and the lower is the probability of the lower payment.

After the first experiment and the second experiment are completed, the computer randomly draws (with the same probability) one of the 10 decision situations. After that, the computer determines your payout from the lottery that you have chosen in this decision situation by a second random draw. For that, the computer uses the probabilities for the higher payment and the lower payment according to the chosen decision situation.

decision	Lottery A	Your decision		Lottery B
		A	B	
1.	2.00 € with 10% or 1.60 € with 90%	o	o	3.85 € with 10% or 0.10 € with 90%
2.	2.00 € with 20% or 1.60 € with 80%	o	o	3.85 € with 20% or 0.10 € with 80%
3.	2.00 € with 30% or 1.60 € with 70%	o	o	3.85 € with 30% or 0.10 € with 70%
4.	2.00 € with 40% or 1.60 € with 60%	o	o	3.85 € with 40% or 0.10 € with 60%
5.	2.00 € with 50% or 1.60 € with 50%	o	o	3.85 € with 50% or 0.10 € with 50%
6.	2.00 € with 60% or 1.60 € with 40%	o	o	3.85 € with 60% or 0.10 € with 40%
7.	2.00 € with 70% or 1.60 € with 30%	o	o	3.85 € with 70% or 0.10 € with 30%
8.	2.00 € with 80% or 1.60 € with 20%	o	o	3.85 € with 80% or 0.10 € with 20%
9.	2.00 € with 90% or 1.60 € with 10%	o	o	3.85 € with 90% or 0.10 € with 10%
10.	2.00 € with 100% or 1.60 € with 0%	o	o	3.85 € with 100% or 0.10 € with 0%

A1.3 Instructions for Main Experiment

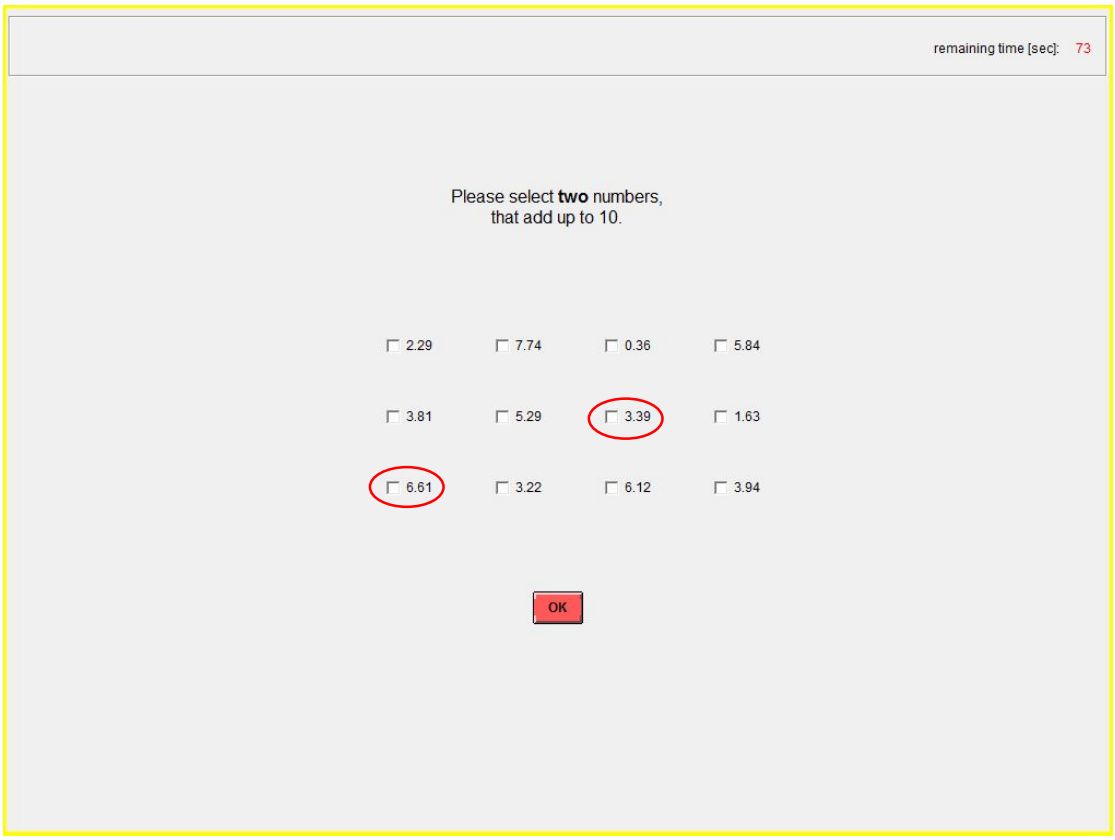
The second experiment consists of two parts. In part 1 of the experiment you have the possibility to earn money. In part 2 of the experiment you complete a tax return over your income.

Part 1

Puzzle Task

The first part of the experiment consists of a math puzzle task. The screen of your computer will display a puzzle with a matrix of 12 numbers. The numbers have two decimal places. These two numbers, that add up to 10, should be found in the matrix. In each matrix with 12 numbers, there are only 2 numbers that exactly add up to 10.

The screen with the matrix looks as follows:



Part 1 of the experiment consists of 6 rounds and in each round you have 3 minutes time to solve as many math puzzles as you like. A maximum of 20 puzzles can be solved in each round. After each round you have a one-minute break before the next round starts.

Your earned income depends on the number of puzzles you solved correctly. For each correctly solved puzzle, you receive an income of 0.42 €. For each incorrectly solved puzzle, you receive an income of 0 €. If you do not solve any puzzle correctly in one round, you earn 0 € in that round. If you solve every puzzle correctly in one round, you earn 8.40 € in that round. After each round, your number of correctly solved puzzles and the resulting income earned in that round will be displayed to you. Your income in each round therefore is as follows:

Earned income in each round =
number of correctly solved puzzles in the corresponding round x 0.42 €

Your total earned income results when adding up the earned income of all six rounds.

After you have earned your pre-tax income (= total earned income) in part 1 of the experiment, you have to complete a tax return on your income in the second part of the experiment. In order to be able to complete the tax return, please note down your earned income after each round on the provided sheet of paper at your place. The sheet of paper serves to help you completing the tax return. Please do not give the sheet of paper back to us, but take it home or dispose of the paper.

Part 2

Tax return

In this study, you have to complete a fictional tax return on your earned income, in order to determine a fictional tax. This means, in part 2 you shall declare the income you earned in part 1. There will be a tax of 25% on the declared income that was stated by you in the tax return. The tax revenue will be used inter alia to finance future research projects at the University of Cologne.

After you have finished part 1 of the experiment, your tax return will be displayed on your screen. The tax return consists of 6 rows, one for the income of each round. At the end of each row you find a field in which you enter the income you want to declare. The income of each round can be already prefilled by the computer. The prefilled values can deviate from your actually earned income. Please check the prefilled values in the tax return. Your declared income can be lower, equal to, or higher than your actually earned income.

To submit your tax return, please press the button "Submit tax return". After submission, you cannot modify your tax return anymore.

Before submitting your tax return, you can also press the button "Compute tax liability" in order to get the total declared income and the resulting tax liability on the basis of your provided information in the tax return displayed. If you want to modify your data, press the button "Modify tax return". You can then change your data in the six income fields. You can press "Compute tax liability" and "Modify tax return" until you are ready with your tax return. As soon as you want to submit your tax return, press the button "Submit tax return".

The tax payable amounts to 25% of your total declared income:

Tax = 0.25 x total declared income

Audit of the tax return

With a probability of 30% your tax return is audited. If you are audited and the declared income of a round does not coincide with your actually earned income of the corresponding round, you have to repay the undeclared tax. Additionally, a fine is charged at the same amount.

Tax repayment = sum of undeclared taxes

Fine = sum of undeclared taxes

The undeclared tax for the income of each round is:

Undeclared tax = 0.25 x (earned income – declared income)

If you are audited and you declared a higher income than you actually earned in that round, you get a tax refund of the overpaid tax. There will be no fine in that case.

Your personal payout of the 2. Experiment

During the 2. Experiment you will receive a personal "bank account", on which your payout relevant amounts will be posted during the experiment. Your account balance will be displayed

to you after each part of the experiment. After part 1, your total earned income of the six rounds will be posted to your account and displayed to you. After you have submitted your tax return, the resulting tax burden of the tax return is posted to your account. If there is an audit and the declared income for each round does not coincide with the actually earned income in the corresponding round, the tax repayment and the resulting fine is debited to the account. If you receive a tax refund, it will be posted to your account.

Your account balance at the end of the 2. experiment is your personal payout of the 2. experiment. Your payout is calculated as follows:

Payout from the 2. experiment = *total earned income in part 1*
- *tax liability from part 2*
- *possible tax repayment*
- *possible fine*
+ *possible tax refund*

Before the 2. experiment begins, you are asked to answer some questions at your computer. The answering of the questions serves to verify your understanding and is not payoff relevant.

A1.4 Comprehension Test

In our comprehension test, subjects have to correctly answer the following questions:

7. How many experiments does the study include?
   Possible answers:
   - o *1*
   - o *2*
   - o *3*
   - o *4*
   - o *More than 4*

8. How is the earned income for each round determined?
   Possible answers:
   - o *The earned income for each round is determined by multiplying the number of correctly solved puzzles in that round with 0.42 €.*
   - o *The earned income for each round is determined by multiplying the total number of correctly solved puzzles in all rounds with 0.42 €.*
   - o *The earned income for each round is determined by multiplying the number of correctly solved puzzles in that round with 0.24 €.*

9. When is a puzzle solved correctly?
   Possible answers:
   - o *When the number 10 is chosen.*
   - o *When so many numbers are chosen, that they sum up to ten.*
   - o *When those two numbers are chosen, that add up to ten.*

10. Which of the following statements regarding the computation of the tax liability is

correct?
Possible answers:
- *The tax amounts to 25% of the total declared income.*
- *The tax amounts to 25% of the total earned income.*

11. With what probability (in percent) will there be an audit of the tax return?
Possible answer: between 0 and 100%

12. How is your personal payout for the second experiment determined when there is no audit?
Possible answers:
- *Payout = total earned income in part 1*
- *Payout = total declared income in part 2 – tax liability in part 2*
- *Payout = fix payout of 4 €*
- *Payout = total earned income in part 1 – tax liability in part 2*

A2 Regressions and further results

Table A2.1: Random Effects Linear Regressions, Within-Subject Comparison (dependent variable: declared share of true income)

	model 1	model 2	model 3	model 4
	blank + correctly prefilled	blank + lower prefilled	blank + higher prefilled	lower prefilled + higher prefilled
prefilled	0.03*	-0.02	0.00	
	(0.02)	(0.02)	(0.02)	
lower prefilled				-0.02
				(0.02)
risk attitude	-0.08	-0.03	-0.01	-0.13**
	(0.06)	(0.05)	(0.05)	(0.05)
female	0.25	0.05	0.32**	0.16
	(0.19)	(0.17)	(0.15)	(0.16)
age	0.01	-0.00	0.02	0.03
	(0.01)	(0.02)	(0.03)	(0.03)
economics	0.04	-0.10	-0.22*	-0.19
	(0.17)	(0.15)	(0.13)	(0.19)
bachelor	0.03	0.22	-0.10	0.39*
	(0.19)	(0.20)	(0.16)	(0.20)
tax experience	-0.26	0.08	-0.16	0.02
	(0.18)	(0.15)	(0.14)	(0.18)
tax knowledge	0.14	-0.05	0.32	-0.11
	(0.24)	(0.26)	(0.21)	(0.27)
tax morality	0.05	0.02	0.03	-0.01
	(0.04)	(0.03)	(0.04)	(0.04)
fairness	-0.01	0.02	0.01	0.03
	(0.03)	(0.03)	(0.03)	(0.04)
decision complexity	-0.05	0.01	0.04	-0.01
	(0.04)	(0.04)	(0.03)	(0.05)
joy	-0.02	-0.05*	-0.01	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
anger	0.03	-0.03	-0.02	0.01
	(0.03)	(0.02)	(0.02)	(0.03)
fear	0.03	0.01	-0.04	-0.06
	(0.04)	(0.06)	(0.03)	(0.04)
guilt	-0.04	-0.05	0.03	-0.00
	(0.05)	(0.06)	(0.05)	(0.05)
income	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
constant	0.51	0.98	0.26	0.33
	(0.55)	(0.87)	(0.74)	(1.06)
no. of observations	235	233	235	220
no. of subjects	40	39	40	37
R-squared				
within	0.0186	0.0052	0.0000	0.0032
between	0.3878	0.3985	0.6197	0.4842
overall	0.3658	0.3641	0.5216	0.4159

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.2: Random Effects Logit Regressions, Within-Subject Comparison (dependent variable: compliance decision)

	model 1	model 2	model 3	model 4
	blank + correctly prefilled	blank + lower prefilled	blank + higher prefilled	lower prefilled + higher prefilled
prefilled	2.29***	-0.28	0.50	
	(0.73)	(0.75)	(0.53)	
lower prefilled				-0.08
				(0.52)
risk attitude	-2.16*	-5.58*	0.49	-1.66**
	(1.16)	(3.07)	(0.64)	(0.65)
female	4.03	9.02	0.99	2.14
	(2.89)	(6.20)	(2.01)	(1.72)
age	0.46**	-1.72	0.29	0.20
	(0.22)	(1.17)	(0.42)	(0.26)
economics	0.67	-3.49	-4.59**	-1.61
	(2.74)	(4.07)	(1.85)	(1.86)
bachelor	1.84	26.45*	-4.27*	5.42**
	(3.01)	(14.63)	(2.38)	(2.47)
tax experience	-8.19***	22.16	-2.80	-0.78
	(2.67)	(14.00)	(1.75)	(1.82)
tax knowledge	3.23	-3.78	4.79*	-0.99
	(3.83)	(6.53)	(2.70)	(2.87)
tax morality	0.85	-0.12	0.23	-0.25
	(0.80)	(0.73)	(0.44)	(0.52)
fairness	-0.56	-0.19	0.33	-0.68
	(0.54)	(0.74)	(0.35)	(0.61)
decision complexity	-1.33*	2.76	-0.05	-0.59
	(0.70)	(2.14)	(0.35)	(0.55)
joy	-0.19	-4.39*	-0.78*	-0.40
	(0.47)	(2.26)	(0.41)	(0.33)
anger	1.28*	-2.03*	-0.90***	0.27
	(0.66)	(1.15)	(0.34)	(0.31)
fear	0.86	-6.05	0.15	-0.42
	(0.63)	(3.88)	(0.44)	(0.44)
guilt	-0.97	3.29	0.33	-1.36**
	(0.79)	(2.64)	(0.57)	(0.63)
income	-0.01	0.03	-0.00	-0.01**
	(0.01)	(0.02)	(0.00)	(0.00)
constant	-4.40	78.45*	-0.79	13.43
	(8.76)	(45.57)	(11.49)	(13.30)
no. of observations	235	233	235	220
no. of subjects	40	39	40	37

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2.3: Consistent Tax Compliance Behavior

	all treatments	treatment B+CP	treatment B+LP	treatment B+HP	treatment LP+HP
always full compliance	45.6%	45.0%	67.5%	35.0%	34.2%
always full evasion	16.5%	20.0%	10.0%	17.5%	18.4%
always full compliance or full evasion	62.0%	65.0%	77.5%	52.5%	52.6%

Note: This table presents an overview of the share of subjects that reveal consistent tax compliance behavior in our experiment. Consistent means that the subjects chose either full compliance or full evasion for all six income fields in their tax return.

Chapter 4

Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups

Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups

Martin Fochmann, Nadja Müller, Nadja Wolf

University of Cologne and University of Hannover

Abstract

Unethical behavior in organizations is usually associated with the risk of negative consequences for the organization and for the involved managers. However, literature on decision making of groups and individuals does either focus on dishonest behavior without the risk of detection and negative monetary consequences or on risk-taking behavior without ethical concerns. Our contribution is that we extend the literature by studying the honesty and risk dimension in one setting. In line with the literature on dishonesty, we observe that groups are less compliant than single individuals. We show that risk concerns have a substantial impact which emphasizes the relevance of the risk dimension when studying honesty behavior in groups. Our results suggest that group interaction changes the norm perception of individuals and increases the willingness to enter in risky non-compliance behavior. Group interaction induces a negative spill-over effect on subsequent individual compliance, thus, the shift in norm perception is persistent even after the group interaction. Our results suggest that the effect of group decision making is stronger when combining the honesty and the risk dimension than what the honesty literature without risk concerns would predict. Furthermore, we show that individuals react differently to group interaction, but differences are systematic and nearly all individuals can be assigned to three types of decision makers.

Keywords

Dishonesty, lying, compliance, risk-taking, group decisions, communication, norms, experiment

JEL-Classification

C91, C92, D03, H26

Acknowledgements: We thank Jochen Hundsdoerfer, Lisa Bruttel, Peter Mohr and participants of the 2018 theem Conference, 2017 arqus Conference, and 2017 GfeW Conference for helpful comments and suggestions.

Chapter 4

Tables

Figures

4.1 Introduction

Decisions in organizations are often taken in groups (e.g., in boards, committees, departments, and teams) rather than by one single individual. Although organizations and their managers should act in a socially responsible, ethically clean way that is compliant with the law, a variety of scandals has damaged this picture. If immoral behavior is disclosed, negative consequences such as penalty payments, compensation claims, loss of reputation, or court hearings can occur for the organization and for the involved managers. In this paper, we study how compliance behavior of groups and individuals differs when non-compliant behavior can have negative consequences. We use a parsimonious design that enables us to identify some of the potential reasons for collective non-compliant behavior.

There is a large literature comparing decision making of groups and individuals with the focus either on immoral behavior (Conrads et al., 2013, Kocher et al., 2017, Sutter, 2009) or on risk-taking behavior (Fahr and Irlenbusch, 2011, Feri et al., 2010, Sutter, 2007). However, there is no overlap of both parts of the literature: dishonest behavior is studied without the risk of negative monetary consequences and risk-taking behavior is studied in situations without ethical concerns. In our setup, non-compliant behavior is both immorally and risky (due to negative consequences when cheating is disclosed). The main contribution of our paper is that we connect both parts of the literature by studying the honesty and risk dimension simultaneously. To be specific, we extend the literature by analyzing compliance differences between groups and individuals where unethical behavior is stochastically audited and penalized.

Unethical behavior in organizations can be disclosed by internal (e.g., by an internal quality management system) or external processes (e.g., by auditors or stakeholders). There are several examples of immoral behavior in organizations leading to negative consequences. A recent example is the Volkswagen's diesel emissions scandal. Volkswagen had to pay huge penalty payments in the US and in Germany, and additionally had to deal with the reputational damage caused by the disclosure of their immoral actions. Further negative monetary consequences arose as the market reacted with a sharp drop in the share price and a massive reduction of diesel car demand. Managers involved in the scandal face the risk of being held accountable for their actions, which leads to serious consequences. They get fired and if they face legal proceedings, large fines or imprisonment can be the consequence. Similar negative consequences resulted for organizations and their managers involved in corruption (e.g., Walmart, Siemens, Halliburton, KBR), financial accounting fraud (e.g., Enron, Worldcom,

Tyco), tax fraud (Lux Leaks, Panama Papers), and insurance fraud (e.g., Hospital Corporation of America, TAP Pharmaceuticals).

In our study, we conduct a controlled laboratory experiment that is based on the design of Kocher et al. (2017). We extend their cheating framework by incorporating a risk dimension: unethical behavior can be disclosed and penalized. We use the following tax compliance context without loss of generalizability. Each individual is an employee of a tax department that is responsible to file the annual income tax return of the organization. Declaring less income than actually earned saves taxes at the organizational level. When non-compliance is disclosed by an audit, the organization has to repay the evaded taxes plus a penalty. In our individual (group) setting, the tax department consists of one employee (three employees). Whereas the single individual decides on the compliance level in the individual setting, the three individuals in the group setting first communicate in an anonymous group chat before voting determines the group decision. Employees receive a fixed remuneration and a variable remuneration that depends on the company's after tax profit. A tax department employee gains from tax evasion if no audit occurs and experiences a disadvantage if an audit occurs. In our setup, we ensure that the monetary payoff for one employee is identical in the individual and group setting.

We use a tax compliance context for different reasons. First, a tax compliance decision implies our two dimensions of interest: the honesty dimension as individuals behave dishonestly when evading taxes (i.e., violating the social norm of paying taxes) and the risk dimension as tax evasion is disclosed with a certain probability and negative monetary consequences result. Second, the tax compliance literature provides us with a standard and well established framework for our analysis that uses the observed compliance level as an indicator for immoral behavior (Allingham and Sandmo, 1972, Torgler, 2002, Hofmann et al., 2008, Alm, 2012). Third, the applied tax compliance framework is a nonstrategic task and ensures that strategic uncertainty about others' behavior does not bias someone's own behavior. This is also why we decided not to implement a public goods experiment with taxes.

Our main results are as follows. First, we observe that compliance is significantly lower in the group than in the individual setting. Consequently, groups are less compliant than single decision makers. This supports the general finding of the literature on dishonest behavior in a setting where unethical behavior can be disclosed and penalized.

Second, we are able to show that risk plays a substantial role when individuals make compliance decisions in groups. Risk arguments are the arguments most frequently mentioned in group communication and, more importantly, the influence of communication on the group's

compliance behavior is mainly driven by arguments relating to risk. This provides strong evidence that risk concerns are non-negligible and have a high impact on dishonest behavior.

Third, our results suggest that communication in groups alters the risk tolerance of the group members. The frequent exchange of risk arguments encouraging non-compliance changes the norm perception of individuals and increases the willingness to enter in risky non-compliance behavior. This is in line with the literature suggesting than learning about preferences of others shifts norm perception (Gino et al., 2009, Kocher et al., 2017) and provides evidence for a conformism effect (Kocher et al., 2013, Lahno and Serra-Garcia, 2015).

Fourth, we find that group interaction induces a negative spill-over effect on subsequent individual compliance. The shift in norm perception through group interaction is a sustaining effect and observable in a significantly lower individual compliance level after group interaction compared to individual compliance before group membership. However, we also observe that compliance is significantly higher in the individual setting after group interaction than in the group setting. This finding suggests that the shift in norm perception is not the only driver for the difference in behavior between groups and individuals in our setup. Otherwise we would have observed the same compliance level in the group setting and in the subsequent individual setting. As the increase in compliance is not observed in a setting exclusively focusing on the honesty dimension (Kocher et al., 2017), our finding suggests that the effect of group membership on behavior is stronger than the honesty literature without risk concerns predicts. This emphasizes the relevance of the risk dimension in group decision making.

Fifth, a categorization of subjects indicates that individuals react differently to group interaction and thus, are heterogeneously in their behavior. However, differences are rather systematic and nearly all individuals can be assigned to three types of decision makers.

The remainder of this paper is structured as follows: In Section 2, we discuss the related literature and develop our hypothesis. Section 3 describes the experimental design. We analyze compliance behavior and treatment differences in section 4. In section 5, we study different types of decision makers and analyze the influence of individual preferences on group compliance. Arguments communicated in the group chats are examined in Section 6. Section 7 concludes.

4.2 Related Literature and Hypothesis

4.2.1 Risk-Taking Behavior

In the 1960s, social psychologists started to investigate the decision behavior of groups and individuals and found a *risky shift* in groups – meaning that groups generally take more risk

than individual decision makers (see Isenberg, 1986, for a literature review). More recent papers, however, observe no differences (Harrison et al., 2013) or even provide evidence for a *cautious shift* implying that groups exhibit more risk aversion than individuals (Masclet et al., 2009, Bolton et al., 2015). Studies using the risk elicitation task of Holt and Laury (2002) often find that groups show both risky and cautious shifts since they are closer to risk-neutrality than individuals (Baker et al., 2008, Shupp and Williams, 2008, He et al., 2012). In particular, groups are more risk-averse in lotteries with a low probability of winning the largest payoff (high risk lottery), but are less risk-averse when this probability is high (low risk lottery). Studies investigating risky investment decisions also fail to provide clear evidence for a risky shift in groups. For example, Bliss et al. (2008) and Bär et al. (2011) observe that team-managed mutual funds are less risky than individually-managed funds. When looking at risk-adjusted performance, Bliss et al. (2008) and Prather and Middleton (2002) find no appreciable differences. In contrast, Rockenbach et al. (2007) observe that groups accumulate more expected value at lower risk. Results are also mixed regarding the level of behavioral biases in risky decisions. Whereas Cheung and Palan (2012) and Sutter (2007) show that behavioral biases are reduced in teams, Whyte (1993) and Rau (2015) observe stronger distortions.

In the literature, three main reasons are discussed why risk-taking can differ between groups and individuals. However, their influence on risk-taking is ambiguous, which could explain why studies fail to find a general tendency. First, there is plenty of evidence that *groups take more rational decisions* than individuals in both strategic and nonstrategic tasks (e.g., Bornstein and Yaniv, 1998, Bornstein et al., 2004, Sutter, 2005, Feri et al., 2010). Reasons are for example that groups are better at learning (Kocher and Sutter, 2005, Cooper and Kagel, 2005, Fahr and Irlenbusch, 2011), reducing behavioral biases (Cheung and Palan, 2012, Sutter, 2007), avoiding extreme decisions (Bär et al., 2011), statistical assessments (Blinder and Morgan, 2005), risk allocation (Rockenbach et al., 2007) and Bayesian updating (Charness et al., 2007). Related to risk-taking decisions, these findings suggest that groups are closer to risk-neutrality than individuals. Indeed, this is what studies, for example applying the Holt/Laury task, generally observe.

The second argument for risk-taking differences between groups and individuals is that *social responsibility leads to conservative risk-taking*. A variety of studies observe that subjects whose risk decisions affect the payoff of all group members (i.e., subjects who are responsible for the group) reveal a lower willingness to take risk (Charness and Jackson, 2009, Reynolds et al., 2009, Ertac and Gurdal, 2012). Bolton et al. (2015) argue that social responsibility can

operate through two channels: “either because decision makers look to avoid blame for bad outcomes or because social responsibility is equated with caution” (p. 110).

Third, *conformism* can cause risk-taking differences. Conformism refers to the phenomenon that individuals change their behavior to match the responses of others (Janis, 1972, Cialdini and Goldstein, 2004, Bolton et al., 2015). In other words, an individual observes what others in the group are doing and adjusts her behavior in that direction. Recent studies robustly observe that individual decisions under risk can be influenced by the risk preferences of peers like other group members (Cooper and Rege, 2011, Kocher et al., 2013, Lahno and Serra-Garcia, 2015). Very related to conformism is *group polarization* which refers to the phenomenon that group decision making is more extreme than the initial tendency of its members (Isenberg, 1986). Suppose for example that the average initial risk-taking tendency of all group members is moderate. Then group discussion might lead to an average shift of group preferences towards a more extreme willingness to take risk since all members adjust their risk preferences above the average group tendency. Please notice that conformism and group polarization can increase or decrease risk-taking in groups compared to individual decision makers depending on whether the initial tendencies of the group members are to be risky or cautious.

4.2.2 Lying and Cheating Behavior

Literature examining unethical behavior of groups and individuals so far focus on settings where lying and cheating behavior involve no risk of being caught and punished. Several studies on lying behavior provide evidence for a *dishonesty shift* in groups – meaning that groups have a stronger inclination to behave unethically than individuals (e.g., Kocher et al., 2017, Conrads et al., 2013, Chytilová and Korbel, 2014, Weisel and Shalvi, 2015, Baeker and Mechtel, 2015). However, not all studies find differences in unethical behavior between groups and individuals (Sutter, 2009, Muehlheusser et al., 2015).

There are four main reasons discussed why groups behave more unethically than individuals. First, groups may have a stronger inclination to behave unethically because *groups behave more strategically* than individuals. Related to the literature on unethical behavior, groups might be more willing to realize the maximum possible gains from cheating than individuals. Several studies find that individuals do cheat for monetary incentives, but not to the full extent as they want to maintain a positive moral self-image (e.g., Mazar et al., 2008, Fischbacher and Föllmi-Heusi, 2013). Second, groups may have a stronger inclination to behave unethically when the *observability of individual actions* within a group is lower, thus

group members might feel less responsible or accountable for their actions (Conrads et al., 2013, Mazar and Aggarwal, 2011).

The third reason is that *communication* within a group can influence unethical behavior. Communication allows group members to exchange arguments in favor of or against unethical behavior. Literature suggests that learning about the preferences of others might change *norm perception.* As outlined before, individuals tend to conform with social norms of their group and change their behavior in order to match the behavior of others. Although a change in norm perception might increase or decrease unethical behavior in groups (depending on the initial norm perception), literature provides clear evidence that learning about arguments in favor of unethical behavior in a group results in more unethical behavior (Gino et al., 2009, Chytilová and Korbel, 2014, Kocher et al., 2017). Further, Kocher et al. (2017) argue that communication enables groups to justify their unethical behavior in different ways than individuals.

Fourth, recent studies suggest that groups may have a stronger inclination to behave unethically when *other people benefit* from dishonest behavior (Gino et al., 2013, Weisel and Shalvi, 2015, Wiltermuth, 2011, Erat and Gneezy 2012, Schweitzer and Hsee, 2002). When a group member's unethical behavior increases not only her own payoff but also the payoff of other group members, this might serve as justification or even a motivation for unethical behavior ("white lie justification"). Others-serving unethical behavior might be judged as less immoral and seen in a more positive way than purely self-serving unethical behavior (Gino et al., 2013, Weisel and Shalvi, 2015).

An argument against a stronger inclination to behave unethical in groups is that unethical behavior in a group can arise *image concerns*. Unethical behavior might be observable by the other group members, which can arise social image (reputational) concerns (Bénabou and Tirole 2006, Dufwenberg and Dufwenberg, 2018, Falk and Tirole, 2016). Furthermore, behaving unethical in a group can increase the saliency of dishonest acts, which might draw attention to one's self-image and one's own honesty standards (Gino et al., 2009, Falk and Tirole, 2016).

4.2.3 Hypothesis

In our study, we combine the honesty dimension with the risk dimension in a tax compliance context. The literature on lying and cheating provides rather clear evidence for a dishonesty shift in groups. Thus, we expect to find a dishonesty shift for group decisions in our experiment. The literature on risk-taking provides mixed results. However, given the three arguments for risk-taking differences discussed in section 2.1, we expect to find a risky shift

for group decisions in our study. First, the risk-taking literature observes that groups take more rational decisions and are closer to risk-neutrality. As our design is – in line with the tax compliance literature – chosen in the way that a risk-neutral and pure money-maximizing decision maker will be non-compliant, we expect that groups are less compliant than individuals according to this argument.

Second, a main argument for a cautious shift in groups is that social responsibility (for other group members) leads to conservative risk-taking. However, social responsibility only plays a minor role in our setting. On the one hand, a subject is only pivotal in the voting process and thus responsible for the payoff of the other group members with a probability of 1/3. On the other hand and more importantly, group members first communicate with each other before each member sends her proposal. Given the high coordination rate of group members observed in Kocher et al. (2017), it is to be expected that each group member sends the proposal agreed on in the communication stage. A single group member is then however not solely responsible for the group outcome. Consequently, we do not expect to observe a cautious shift resulting from social responsibility in our study.

Third, we expect that conformism is a main driver in our group setting. As we allow subjects to communicate in a group chat, subjects learn about the preferences of the other group members and conformity effects might shift individual preferences towards these preferences. However, this depends on the disclosed preferences of the others. If we assume a high frequency of arguments in favor of non-compliance in our group chats (in line with the finding of Kocher et al., 2017, that arguments for dishonesty are mentioned more frequently than arguments for honesty), we expect that conformity effects will lead to less compliant behavior in groups. Consequently, the compliance level will be lower in the group than in the individual setting.

Combining the honesty and the risk dimension, we expect that groups are more non-compliant than individuals. Furthermore, by adding up both dimensions, we expect that the difference in behavior between groups and individuals increases in size. In other words, groups have a stronger inclination to behave non-compliant than individuals and this effect is stronger than the sole inclination to behave unethically or the sole inclination to take more risks. Therefore, we formulate the following research hypothesis:

Hypothesis: *Compliance is lower in the group than in the individual setting.*

4.3 Experimental Design and Sample

4.3.1 Decision Task and Payoff Functions

In our experiment, each subject faces the decision of an employee who has to declare the income of her company for tax purposes in a tax return. The actual income of the company, which is known by the employee, is fixed and amounts to 1,000 Lab-points.[31] The employee decides how much of the actual income is reported in the company's tax return where all integer values from 0 to 1,000 are allowed. The company has to pay a corporate tax which is 25% of the reported income.[32] With a probability of 30%, the reported income is audited. If an audit reveals that the reported income is less than the actual income of 1,000, the company has to pay a penalty that is twice the evaded tax. In fact, company has to repay the evaded tax plus a fine which is set to the tax amount evaded for reason of simplification. If the company is not caught misreporting, no consequences occur. The company's after tax profit is therefore

Company's after tax profit = 1,000 – 25% · reported income

if no audit occurs, and is as follows if an audit occurs:

Company's after tax profit = 1,000 – 25% · reported income
– 2 · 25% · (1,000 – reported income).

The employee's payoff (and thus participant's payoff from the experiment) is determined by a fixed remuneration of 20 Lab-points and a variable remuneration which amounts to 20% of the company's after tax profit. Consequently, the reporting decision of an employee affects her payoff since it affects the after tax profit of the company. The employee's payoff function can be written as follows:

Employee's payoff = 20 + 20% · company's after tax profit

4.3.2 Individual and Group Setting

We use an individual and a group setting to address our research questions. In the individual setting, each individual makes an own reporting decision independently of other participants of the experiment. In the group setting, three participants are randomly assigned to one group. The task of the group is the same as in the individual setting (i.e., reporting the company's income). Each group member decides individually about how much income should

[31] To abstract and simplify monetary values, we use Lab-points as currency units, where 1 Lab-point exactly corresponds to EUR 0.10.

[32] Our chosen levels for the tax rate, audit probability, and penalty are similar to those used in other tax compliance experiments (see Alm et al., 1995, Andreoni et al., 1998, Torgler, 2002, Hofmann et al., 2008, and Alm, 2012 for excellent literature reviews).

be reported by the group. The median of the three proposals determines the income reported by the group. Group members only see the voting outcome, but not the individual proposals of the other group members. Before the individual decision is made, subjects are allowed to communicate with the other members of their group by sending text messages with an anonymous chat presented on the computer screen. This in the only difference to the individual setting. All other design parameters remain unchanged. Most importantly, the calculations of the company's after tax profit and employee's payoff are the same in both settings. Thus, each subject in the individual and group setting receives a fixed remuneration of 20 Lab-points and a variable remuneration of 20% of the company's after tax profit.

To compare the behavior between the individual and group setting, we adapt the experimental design presented by Kocher et al. (2017). In total, we have three treatments. Each treatment consists of three parts and each part consists of three consecutive income reporting decisions (i.e., nine decision situations in total). Income reporting decisions neither vary between the three parts nor between the three decisions within one part. The only difference is that in a particular part either the individual or the group setting is applied. In treatment I-I-I, the individual (I) setting is applied in all three parts. In treatment G-G-G, the group (G) setting is applied in all three parts. In treatment I-G-I, the individual setting is applied in the first part, the group setting in the second, and the individual setting again in the third part. A subject participates only in one of the three treatments. Figure 1 presents our experimental design and highlights the treatment differences.

Our approach enables us to analyze the differences between individual and group setting in different ways. Comparing treatment I-I-I with G-G-G allows a between-subject analysis. Comparing the three parts of treatment I-G-I allows a within-subject analysis. As we have three income reporting decisions per part, we are able to analyze behavior over time (e.g., learning effects) within each part and across different parts.

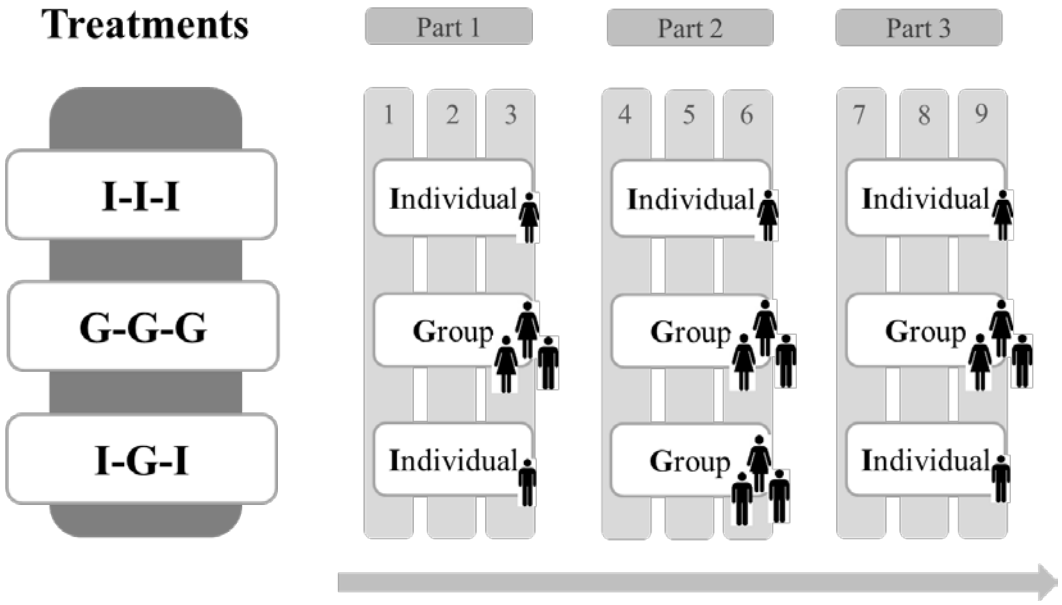


Figure 1: Experimental Design

Note: This figure highlights the differences among the treatments.

4.3.3 Experimental Protocol

At the beginning of each part, participants receive written instructions in which all part-related information are presented. The instructions are available in our appendix A1. In the instructions for the first part, subjects are informed that the entire experiment consists of three parts in total and that each part consists of three decisions. Furthermore, participants are informed that, at the end of the experiment, for each subject one out of the nine decision situations is chosen randomly to determine the individual payoff.

Before the actual experiment is executed, we measure subjects' willingness to take risk with the Holt and Laury (2002) task (in EUR). We use the total number of high risk lottery choices (out of 10) as our measure. Consequently, subject's willingness to take risk is measured on an 11-point Likert scale, where 0 = not willing to take risk and 10 = highly willing to take risk. The amount earned in the lottery task is also paid out to each participant at the end of the experiment. We obtain further information about the individual characteristics of our participants (e.g., gender, age, tax morale, etc.) in a questionnaire. At the end of the experiment, each participant receives her total payoff from the experiment plus a show-up fee of EUR 4 in cash.

In the group setting, three subjects are randomly assigned to one group. This assignment was invariably kept for the rest of the experiment when the group setting is applied. This implies for the treatment G-G-G that a subject is always confronted with the same group members in all three parts. In the treatment I-G-I, a subject is only confronted with the same group members

in the three decision situations of the second part. In the group setting, group members can communicate with an anonymous real-time chat. This chat enables group members to communicate in a free-form by entering text in an input mask. It is ensured that the identity of a group member is not revealed and the possibility of side-payments is excluded. All messages sent in the chat are seen by all group members. Thus, no message can be sent to only one group member. Each group member can unanimously decide to leave the chat. The number of messages sent is not restricted, but the chat automatically ends after five minutes.

Although we use a simple setting, each participant receives a pocket calculator and a computerized "what if" calculator for her own calculations. The latter allows subjects to automatically calculate, for example, the company's after tax profit and subject's payoff for the situation with or without an audit. In the individual and group setting, the "what if" calculator is displayed when subjects decide individually on the reported income. In the group stage, the calculator is further displayed in the chat stage.

All information presented here are provided to the subjects in the instructions. The experimental software was programmed and conducted with the software z-Tree (Fischbacher, 2007). Participants were recruited with ORSEE (Greiner, 2004).

4.3.4 Sample and Data

The experiment was conducted at the computerized experimental laboratory of the University of Cologne (CLER) in March and April 2017. In total, 189 subjects (mainly undergraduate students, 97 females and 92 males) participated and earned, on average, EUR 24.46 in approximately 105 minutes (approximately EUR 13.98 per hour). Table 1 gives an overview of all our variables and presents some descriptive results.

Table 1: Overview of Variables

Variable	description	mean
declared income	income declared in tax return (0 to 1000)	
treatment I-G-I	Individual-Group-Individual	
treatment G-G-G	Group-Group-Group	
treatment I-I-I	Individual- Individual- Individual	
part	1; 2; 3	
period	1; 2; 3 in each part	
last period audit	audit in previous period = 1; elsewise = 0	
Ex-post questionnaire		
female	female = 1; male = 0	51.32%
risk attitude	Holt&Laury (2002) risk measure	4.21 / 10
age	in years (18 to 66)	24.95
economics	study with more than one lecture in economics = 1; elsewise = 0	57.14%
bachelor	study with a bachelor's degree =1, elsewise = 0	57.14%
tax experience	experience with tax returns =1, elsewise = 0	41.27%
tax knowledge	tax knowledge = 1; no tax knowledge = 0	14.29%
tax morality	0 to 9; low tax morality = 0; high tax morality = 9	6.87
positive reciprocity	0 to 10; low positive reciprocity = 0; high positive reciprocity = 10	8.22
negative reciprocity	0 to 10; low negative reciprocity = 0; high negative reciprocity = 10	5.46
fairness	0 to 10; low perceived fairness of tax and control system in experiment = 0; high perceived fairness of tax and control system in experiment = 10	6.41
decision complexity	0 to 10; low perceived decision complexity in experiment = 0; high perceived decision complexity in experiment = 10	1.72
joy	0 to 10; felt no joy during experiment = 0; felt high joy during experiment = 10	6.17
anger	0 to 10; felt no anger during experiment = 0; felt high anger during experiment = 10	3.81
fear	0 to 10; felt no fear during experiment = 0; felt high fear during experiment = 10	1.86
guilt	0 to 10; felt no guilt during experiment = 0; felt high guilt during experiment = 10	1.57
income	in Euro (monthly income after fixed costs)	324.10
religious	praying at least once a week = 1; elsewise = 0	22.22%
what if calculations	number of what if calculations used before submitting declared income	0.85

Note: This table presents all variables of our experiment.

4.4 Results: Treatment Differences

Our compliance measure is the income declared by a subject in a period. As the actual income was kept constant across periods and treatments (1,000), we can use the absolute values of declared income for our treatment comparison. Please notice that in case of a group decision, each subject decides individually about how much income should be declared by the group. The group median of these proposals then determines how much income is declared by the group. As we are interested in how individual willingness to report income truthfully varies across treatments, we use the income each subject declared in the following analyses (if not

stated differently). Please notice that using the group median instead would lead to the same findings and conclusions.

Figure 2 shows the declared income on average for each treatment. In part 1, the mean declared income is 463 in treatment I-I-I, 291 in G-G-G, and 392 in I-G-I. The difference between the treatments I-I-I and I-G-I is not statistically significant (Mann-Whitney U-test, p-value > 0.1, two-tailed).[33] This was to be expected because the setting (i.e., individual setting) is completely identical in part 1 in both treatments. In contrast, the difference between treatment G-G-G and each of the other two treatments is statistically significant (p-value < 0.01 for both comparisons). Consequently, in the group setting we observe lower compliance than in the individual setting in part 1. The same finding results if we compare the declared income observed in part 2 between treatment I-I-I (462) and treatment I-G-I (162) (p-value < 0.01) and if we compare the declared income observed in part 2, part 3, or parts 1 to 3 between treatment I-I-I (part 2: 462, part 3: 479, parts 1 to 3: 468) and treatment G-G-G (part 2: 238, part 3: 226, parts 1 to 3: 252) (p-value < 0.01 for all comparisons). Consequently, declared income is always lower in the group than in the individual setting. Thus, our hypothesis is supported.

Result 1: *The compliance level is significantly lower in the group than in the individual setting.*

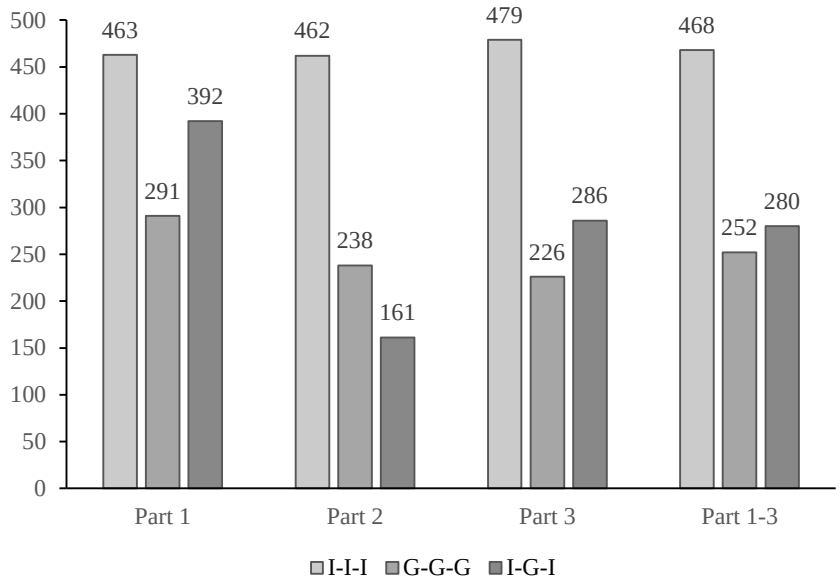


Figure 2: Declared Income

Note: This figure shows the declared income for each treatment for the different parts on average.

[33] Please notice that our results remain robust if we use the mean income declared by a subject in one part as dependent variable (i.e., only one observation per subject and part).

The only exception occurs if we compare the declared income observed in part 3 between treatment I-G-I (286) and treatment G-G-G (226). Although the compliance level is higher in treatment I-G-I, the difference is not statistically significant anymore (p-value > 0.1). In contrast, we observe significantly lower declared income when comparing treatment I-G-I (286) and treatment I-I-I (479) in part 3 (p-value < 0.01). This provides evidence for a spill-over effect in treatment I-G-I. Whereas declared income is relatively high in the first individual setting (part 1: 392), it decreases markedly in the group setting (part 2: 162), and increases (but less markedly) in the second individual setting (part 3: 286). The differences between all three parts are statistically significant in treatment I-G-I (p-value < 0.01 for all three comparisons). Consequently, after the decrease from part 1 to part 2, individuals increase compliance from part 2 to part 3 in treatment I-G-I. However, compliance is significantly lower than observed in part 1.

Result 2: *Group interaction induces a negative spill-over effect on subsequent individual compliance: compliance is lower after a group interaction than before.*

Figure 3 displays the histograms for treatments I-I-I and G-G-G (pooled over all parts)[34] and Figure 4 shows the histograms for each part of treatment I-G-I. As standardly observable in tax compliance experiments, we observe that a relatively high number of subjects chose either to report their income truthfully or to report zero income. For example, an income of zero (1,000) was declared in 36% (30%) of all our observations in treatment I-I-I. Furthermore, we observe peaks for each of the one hundred values (i.e., 100, 200, etc.) in all treatments. When comparing the individual and group settings, we find a general shift in distribution (i.e., reported income is generally decreased in the group setting). This also has the consequence that the fraction of zero (1,000) income reports is much higher (lower) in case of the group setting (treatment G-G-G, part 2 of treatment I-G-I) than in case of the individual setting (treatment I-I-I, part 1 and 3 of treatment I-G-I). Beyond this general shift, we do not find any important difference between the distributions. Thus, the histograms confirm our previous results.

[34] Please notice that we did not find any meaningful difference between the histograms when separating the three parts within treatment I-I-I or G-G-G. Thus, we decided to not report the histogram for each part of treatment I-I-I and G-G-G.

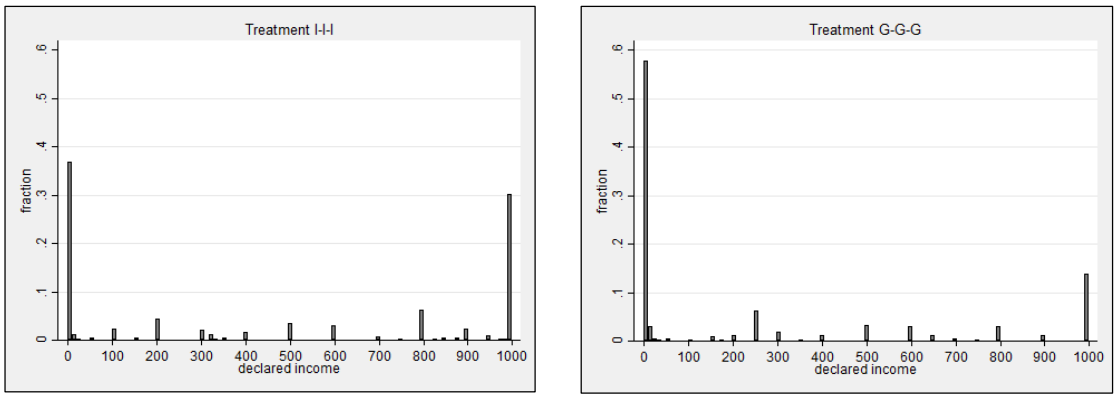


Figure 3: Histograms for Treatments I-I-I and G-G-G (data pooled over all parts)

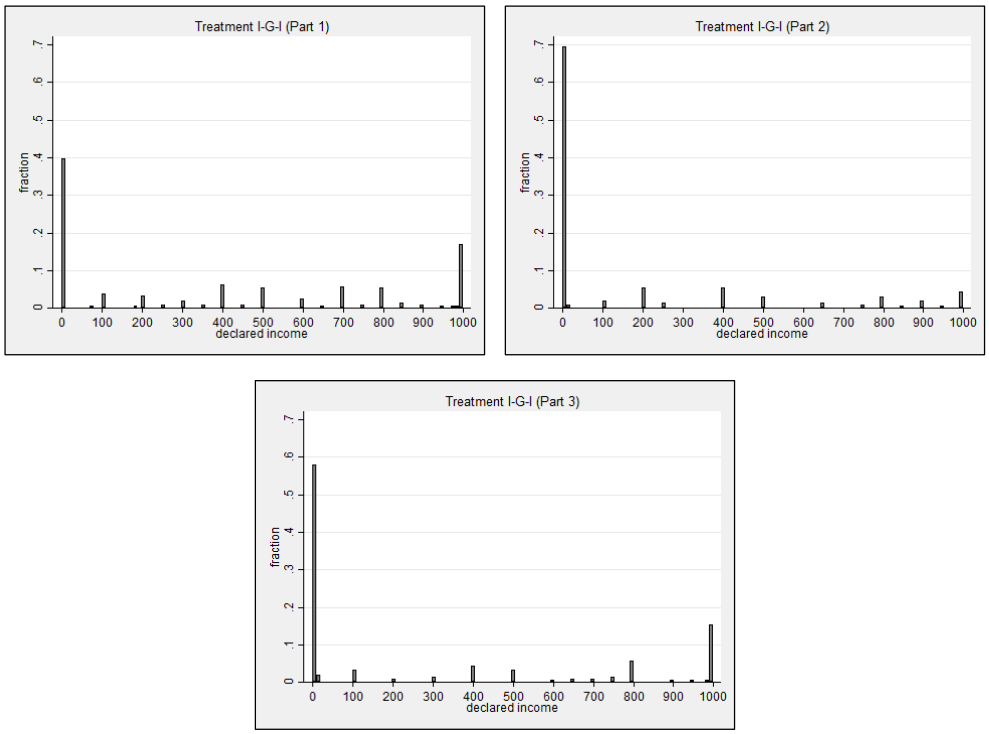


Figure 4: Histograms for Each Part of Treatment I-G-I

In our group setting, each subject decides individually about how much income should be declared. However, this decision is very likely not independent of the other group members. Furthermore, an individual declares income repeatedly in our experiment and therefore the decisions within one subject are not independent. Consequently, the assumption of the nonparametric Mann-Whitney U-test that observations must be independent is not fulfilled. To corroborate our descriptive and nonparametric results, we therefore run linear regressions. We use the income declared by each subject in each period as the dependent variable. As subjects face repeated decisions over several periods and in the group setting face repeated decisions within one group, we run multi-level mixed effects linear regressions to capture more levels of

dependence.[35] To account for heterogeneity across individuals and across groups, subject-specific effects, group-specific effects, and the conventional equation error term are included in the estimated equations. Consequently, this allows us to cluster at the group and at the individual level.

Table 2 reports the outcome for the comparison of treatment I-I-I and G-G-G (regression coefficients, standard errors in parentheses). In model 1, we only regress on the treatment dummy "Treatment G-G-G". Since the treatment I-I-I serves as default, the coefficient of the treatment dummy measures the difference between treatment I-I-I and G-G-G. We observe a significant decrease of declared income in treatment G-G-G. Thus, we find a lower compliance level in the group than in the individual setting and are therefore able to confirm our stated result 1 by this analysis.

To control for differences between our three parts, we additionally regress on the dummies "Part 2" and "Part 3" (which take the value of 1 if the decision was made in the respective part, 0 otherwise) in model 2. Coefficients of the interaction terms "Part 2 X Treatment G-G-G" and "Part 3 X Treatment G-G-G" measure the difference between treatment I-I-I and G-G-G in part 2 and 3, respectively. Statistical significance between our two part dummies and our two interaction terms was checked by Wald tests, and the resulting p-values are reported at the bottom of the table. Again we observe a negative and significant effect of the treatment dummy as in model 1, but do not find any significant effect for the additionally included variables. The only exception is observed for the interaction term "Part 3 X Treatment G-G-G". The coefficient is negative and significant at the 5%-level. This implies that in addition to the (negative) main treatment effect, declared income is even further decreased in the third part of treatment G-G-G compared to treatment I-I-I. This is supported by our graphical analysis. In figure 2, we show that reported income decreases over the three parts in treatment G-G-G whereas it is almost constant over the three parts in treatment I-I-I. Thus, we can conclude that reported income is generally lower in the group than in the individual setting and that this effect is even more pronounced in the third part of the experiment.

In model 3 and 4, we use the same specifications as in model 1 and 2, but further include the dummy variable "last period audit" (which takes the value of 1 if income declaration was audited in the previous period, 0 otherwise) and "period (1 to 3) within part" (which denotes in which period within a respective part the decision was made, values from 1 to 3) as well as individual specific variables like gender, age, etc. We incorporate all 19 individual variables reported in table 1. We report the coefficient of the dummy variable "female subject" in table

[35] A detailed description of multi-level modelling is, for example, given by Moffatt (2015).

2 (which takes the value of 1 if the decision was made by a female person, 0 otherwise). All other individual variables are not displayed. However, the complete set of all regression results is presented in our appendix A2. Again, we observe the same results as in the previous models 1 and 2. In line with the literature on tax compliance, we observe that women are significantly more compliant than men and that individuals are significantly less compliant if they were audited in the previous period.[36]

Table 2: Treatment I-I-I vs. G-G-G: Multi-Level Mixed Effects Linear Regressions (dependent variable: declared income)

	Treatment I-I-I vs. G-G-G			
	model 1	model 2	model 3	model 4
Treatment G-G-G	-216.60**	-172.20**	-224.38***	-176.22**
	(84.81)	(87.06)	(83.39)	(87.73)
Part 2		-0.60		29.98
		(26.36)		(29.17)
Part 2 X Treatment G-G-G		-51.98		-46.38
		(34.03)		(36.48)
Part 3		16.03		49.13
		(26.36)		(30.10)
Part 3 X Treatment G-G-G		-81.22**		-71.61*
		(34.03)		(37.18)
last period audit			-60.15***	-59.92***
			(20.31)	(20.17)
period (1 to 3) within part			7.56	8.50
			(8.74)	(9.11)
female subject			154.66***	153.82***
			(42.40)	(42.42)
other individual controls	NO	NO	YES	YES
constant	468.15***	463.01***	319.62**	284.46*
	(50.72)	(52.95)	(152.17)	(154.55)
no. of observations	1,080	1,080	960	960
no. of subjects	120	120	120	120
no. of independent groups	72	72	72	72
Wald test:				
Part 2 = Part 3		p = 0.5281		p = 0.4401
Part 2 X Treatment G-G-G = Part 3 X Treatment G-G-G		p = 0.3902		p = 0.4309

Note: In this table, the results of multi-level mixed effects linear regressions are presented with declared income as dependent variable (regression coefficients, standard errors in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

In table 3, we use the same approach and specifications to analyze the differences between treatments I-I-I and I-G-I. In models 5 and 7, we observe a significantly lower compliance level in treatment I-G-I. However, models 6 and 8 reveal that the treatment difference is only

[36] The last result is in line with the "bomb crater effect" first observed by Mittone (2006) and further analyzed by, for example, Maciejovsky et al. (2007) and Kastlunger et al. (2009). This effect describes the tendency of subjects to decrease their compliance rates immediately after they are audited.

significant in part 2 and 3. This is indicated by the negative and significant coefficients of both interaction terms “Part 2 X Treatment I-G-I” and “Part 3 X Treatment I-G-I”. In contrast, no significant difference is observed in part 1 where the individual setting is applied in both treatments (indicated by the non-significant dummy “Treatment I-G-I” in models 6 and 8). Consequently, compliance is lower in the group than in the individual setting (result 1).

In both models 6 and 8, we find that compliance increases from part 2 to part 3 in treatment I-G-I. This is indicated by the higher (i.e., less negative) coefficient of the interaction term “Part 3 X Treatment I-G-I” than of “Part 2 X Treatment I-G-I”. Wald tests reveal that both coefficients differ significantly (see last row for the corresponding p-values). However, this increase does not compensate the large difference between both treatments occurred in part 2. Consequently, in part 3 compliance is still significantly lower in treatment I-G-I than in I-I-I (indicated by the negative and significant coefficient of the interaction term “Part 3 X Treatment I-G-I”). This provides further evidence for the discussed spill-over effect in treatment I-G-I (result 2). In line with the regression results presented in table 2, we find a positive and significant effect of “female subject” and a negative and significant effect of “last period audit”.

We can summarize that our descriptive and nonparametric findings are supported by our multi-level mixed effects linear regressions that account for more levels of dependence (i.e., subject and group level). In particular, we are able to confirm result 1 that compliance is lower in case of the group setting than in case of the individual setting. Furthermore, we find support for result 2 that group interaction (part 2) causes a negative spill-over effect on individual compliance in treatment I-G-I. The resulting consequence is that individual compliance is lower after a group interaction (part 3) than before (part 1).

Table 3: Treatment I-I-I vs. I-G-I: Multi-Level Mixed Effects Linear Regressions (dependent variable: declared income)

	Treatment I-I-I vs. I-G-I			
	model 5	model 6	model 7	model 8
Treatment I-G-I	-188.49**	-71.16	-187.90**	-66.25
	(73.76)	(77.15)	(79.06)	(83.33)
Part 2		-0.60		35.20
		(30.10)		(32.16)
Part 2 X Treatment I-G-I		-229.76***		-217.66***
		(39.19)		(40.88)
Part 3		16.03		53.01
		(30.10)		(32.46)
Part 3 X Treatment I-G-I		-122.23***		-107.36***
		(39.19)		(40.82)
last period audit			-42.16**	-57.21***
			(18.82)	(18.21)
period (1 to 3)			22.17**	13.75
			(10.27)	(10.41)
female subject			178.48***	177.84***
			(49.32)	(49.07)
other individual controls	NO	NO	YES	YES
constant	468.15***	463.01***	342.57	361.21
	(47.39)	(50.48)	(222.89)	(224.33)
no. of observations	1053	1053	936	936
no. of subjects	117	117	117	117
no. of independent groups	71	71	71	71
Wald test:				
Part 2 = Part 3		p = 0.5806		p = 0.52492
Part 2 X Treatment I-G-I = Part 3 X Treatment I-G-I		p = 0.0061		p = 0.0026

Note: In this table, the results of multi-level mixed effects linear regressions are presented with declared income as dependent variable (regression coefficients, standard errors in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

4.5 Treatment I-G-I: Types of Decision Makers and Group Composition

4.5.1 Types of Decision Makers in Treatment I-G-I

The design of treatment I-G-I allows us to study an individual's compliance behavior before group interaction (part 1), within the group (part 2) and after group interaction (part 3). The research question is whether we find different patterns of compliance behavior that allow us to identify different types of decision makers. For categorization, we use the individual's mean declared income in each part. We observe eight (out of nine possible) different types. However, 89% of all subjects can be assigned to three types. First, we find subjects revealing a constant reporting behavior in all three parts. These subjects are classified as C-Type ("Constant-Type"). Second, we find subjects whose individual compliance is significantly

higher in the individual setting (part 1), then decreases in the group setting (part 2), and significantly increases again in the last individual setting (part 3). These subjects are classified as V-Type. Third, we find subjects whose individual compliance is significantly higher in the first individual setting, then decreases in the group setting, but then – in contrast to the V-Type – stays at this lower compliance level in the last individual setting after the group interaction. These subjects are classified as L-Type.

Figures 5 to 7 display examples for each of these three decision maker types. To account for slight fluctuations in the compliance level, we allow for a tolerance level of 100 Lab-points.[37] This means, we only recognize a change in behavior if the difference in declared income between two parts is at least 100 Lab-points. A difference below this threshold is interpreted as a constant behavior between the two parts.

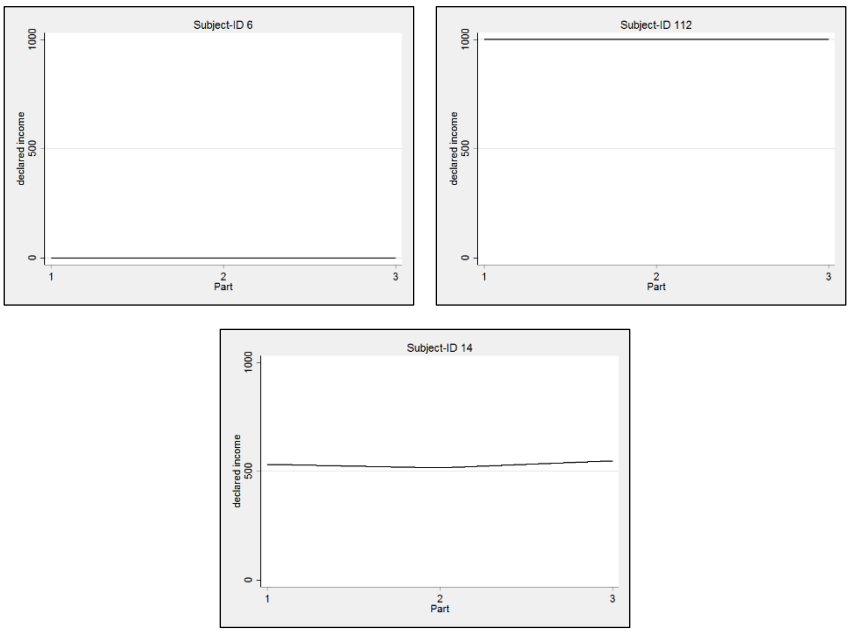


Figure 5: Examples of C-Types

[37] As robustness tests, we also used a tolerance level of 150 and 200 Lab-points. All results are robust to this variation.

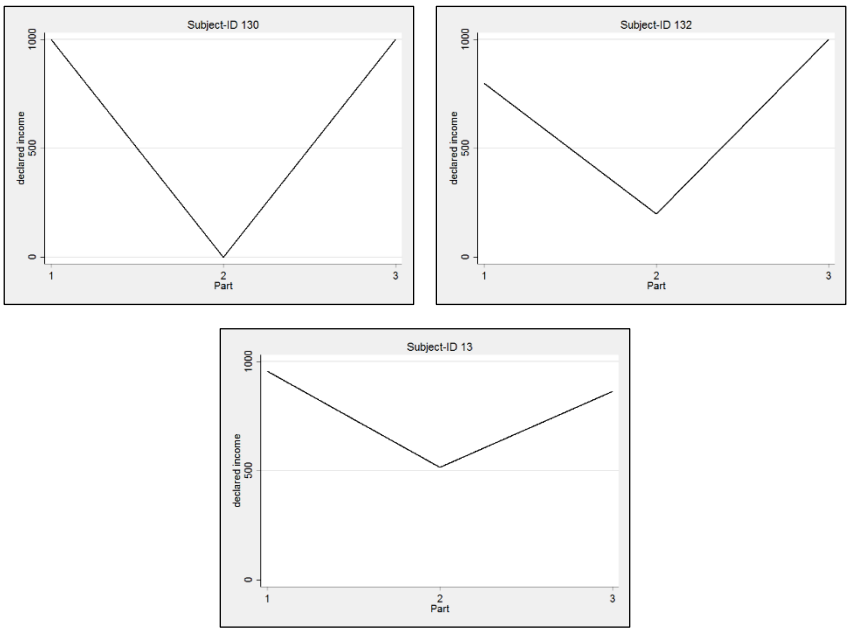


Figure 6: Examples of V-Types

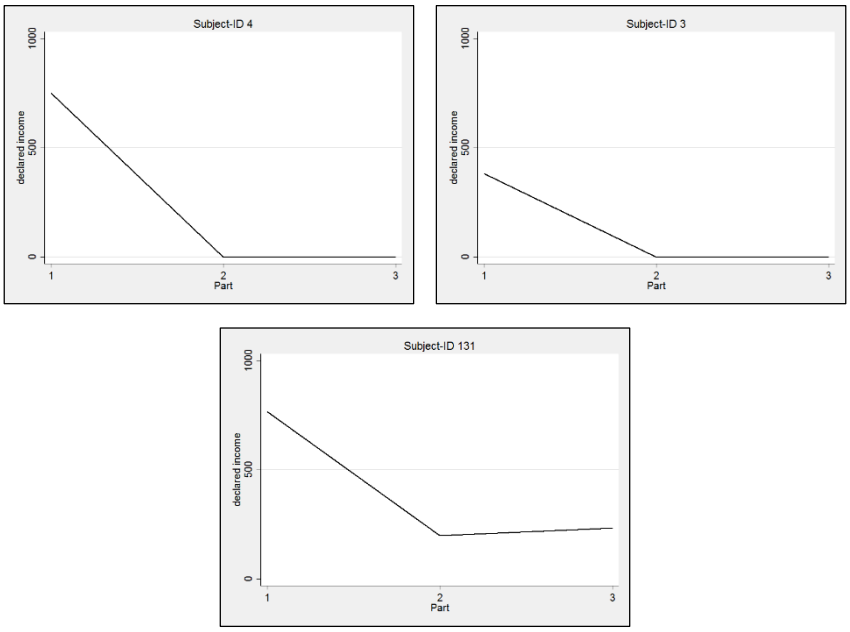


Figure 7: Examples of L-Types

The distribution of subjects in our categorization of different decision maker types is displayed in Table 4. As mentioned before, 89% of all subjects can be assigned to the three types mentioned above. 38% of subjects are classified as C-Types and thus reveal an almost constant behavior across all three parts. In total, 54% of all subjects reported a higher income in the first individual setting than in the group setting (37 out of 69). 41% of these subjects (15 out of 37) increased their reported income in part 3 (V-Type, 22% of all subjects) and 54% (20 out of 37) stayed at the low group level in part 3 (L-Type, 29% of all subjects). There are no

significant differences between the three types regarding revealed risk preferences and stated tax morale level (Kruskal-Wallis test, p-values above 0.1, two-tailed).

Interestingly, we observe substantial gender differences. 54% of all men are categorized as C-Types whereas only 21% of all women refer to this category. In contrast to that, a much higher share of women is categorized as V-Types (32% of all women), whereas only 11% of all men belong to this category. Having a closer look at the compliance level for the C-Types, we find that men revealing a constant behavior mostly declare zero income (84% of all male C-Types), but only 57% of the female C-Types do so. In contrast, 29% of all female C-Types declare their income truthfully (i.e., 1,000) whereas no male C-Types do so.

Result 3: *Almost all subjects can be assigned to three types of decision makers: C-Type (38%), V-Type (22%), and L-Type (29%). Male subjects are mainly C-Types and the V-Type category mainly consists of female subjects.*

Furthermore, we observe that 54% of all subjects show a significant lower compliance level in the group setting than in the first individual setting. For these subjects, we again find substantial gender differences for the individual compliance behavior following the group interaction. 48% of the women (11 out of 23) increase their compliance behavior afterwards and 48% stay on the lower compliance level as in the group setting, i.e. they show a negative spill-over effect. For the male subjects we find that 64% (9 out of 14) show the negative spill-over effect, and only 29% increase their individual compliance again after group interaction. Consequently, our results suggest that the negative spill-over effect of group interaction on subsequent individual compliance (see result 2) is less pronounced for women than for men.

Table 4: Types of Decision Makers in Treatment I-G-I

	all subjects (69 subjects)				men only (35 subjects)				women only (34 subjects)			
	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
C-Type (“—”)	**26**	**38%**	**26**	**38%**	**19**	**54%**	**19**	**54%**	**7**	**21%**	**7**	**21%**
V-Type (“∨”)			**15**	**22%**			**4**	**11%**			**11**	**32%**
L-Type (“└—”)	**37**	**54%**	**20**	**29%**	**14**	**40%**	**9**	**26%**	**23**	**68%**	**11**	**32%**
“＼”-Type			2	3%			1	3%			1	3%
“／”-Type			1	1%			1	3%			0	0%
“∧”-Type	5	7%	1	1%	2	6%	0	0%	3	9%	1	3%
“⌒”-Type			3	4%			1	3%			2	6%
“⌒／”-Type	1	1%	1	1%	0	0%	0	0%	1	1%	1	1%
“⌒＼”-Type			0	0%			0	0%			0	0%

Note: This table presents the types of decision makers categorized in treatment I-G-I.

4.5.2 Group Composition in Treatment I-G-I

In this section we want to examine whether the outcome of a group in treatment I-G-I depends on the group members' individual compliance levels, individual risk preferences or sexes.

The influence of individual compliance on group behavior

First, we have a closer look at the influence of individual compliance and analyze whether the group consists of rather compliant or non-compliant group members. To categorize each participant as a more or less compliant subject we use the income declared by each subject in the first individual setting of treatment I-G-I (part 1). In particular, an individual is categorized as a more (less) compliant subject if her declared income in each period (out of three periods) is equal to or greater than (less than) a certain threshold.[38] We vary the threshold from 100 to 1,000 Lab-points. Whereas 1,000 Lab-points implies full compliance, the median in the first part of treatment I-G-I was 300-Lab-points. A group is called a more (less) compliant subjects dominated group if the group consists of at least two more (less) compliant subjects. Table 5 shows the income finally declared by the group on average for each of the two group categories (Panel A) and for groups separated by the number of more compliant subjects in the group (Panel B). The corresponding number of groups is presented in parentheses.

We observe that more compliant subjects dominated groups reveal a significantly higher compliance level than less compliant subjects dominated groups. This result is independent of the applied threshold. For example, when using the median as the threshold (i.e., 300 Lab-points), we find that the nine more compliant subjects dominated groups declared an income of 322.2 Lab-points on average in the group setting of treatment I-G-I whereas the fourteen less compliant subjects dominated groups declared an income of only 45.2 Lab-points. A similar result is observed when separating groups by the number of more compliant subjects in the group. When using the 300-threshold again, we find that groups with zero more compliant subjects declared an income of zero Lab-points, with one 70.4 Lab-points, and with two 322.2 Lab-points on average. All differences are significant. Interestingly, we robustly observe that even one more compliant subject in a group leads to a higher compliance level of the entire group. However, the strongest effect is found when the group consists of two (or three) more compliant subjects. Consequently, we can summarize that income declared by a group increases

[38] Please notice that we also ran the analysis by using a slight modified version of this categorization. In particular, an individual is categorized as a more (less) compliant subject if the declared income in at least two out of the three periods is equal to or greater than (less than) a certain threshold. All results are robust to this variation.

with the number of (rather) honest group members and therefore depends on individual compliance preferences of the group members (measured before group interaction).

Table 5: Finally Declared Income on Average in the Group Setting of Treatment I-G-I

	Threshold for categorization of more/less compliant subjects									
	100	200	**300**	400	500	600	700	800	900	1000
Panel A: more vs. less compliant subjects dominated groups										
more compliant subjects dominated groups	223.1 (13)	290.0 (10)	**322.2 (9)**	362.5 (8)	594.4 (3)	466.7 (2)	466.7 (2)	433.3 (1)	--- ---	--- ---
less compliant subjects dominated groups	63.3 (10)	48.7 (13)	**45.2 (14)**	42.2 (15)	87.5 (20)	123.8 (21)	123.8 (21)	140.9 (22)	153.6 (23)	153.6 (23)
MWU-test (p-value, 2-tailed)	0.018	< 0.001	**< 0.001**	< 0.001	< 0.001	0.002	0.002	0.012	---	---
Panel B: number of more compliant subjects in group										
three	393.3 (5)	683.3 (2)	**--- ---**	--- ---	--- ---	--- ---	--- ---	--- ---	--- ---	--- ---
two	116.7 (8)	191.7 (8)	**322.2 (9)**	362.5 (8)	594.4 (3)	466.7 (2)	466.7 (2)	433.3 (1)	--- ---	--- ---
one	105.6 (6)	79.2 (8)	**70.4 (9)**	59.3 (9)	137.5 (12)	250.0 (10)	312.5 (8)	377.78 (6)	383.3 (6)	356.7 (5)
zero	0 (4)	0 (5)	**0 (5)**	16.7 (6)	12.5 (8)	9.1 (11)	7.7 (13)	52.1 (16)	72.5 (17)	97.2 (18)
MWU-test (p-value, 2-tailed)										
three vs. two	0.004	0.007	**---**	---	---	---	---	---	---	---
two vs. one	0.721	0.050	**0.001**	< 0.001	< 0.001	0.088	0.227	---	---	---
one vs. zero	0.050	0.062	**0.080**	0.407	0.090	< 0.001	< 0.001	< 0.001	0.001	0.008

Note: This table presents the income finally declared by groups on average in the second part of treatment I-G-I dependent on the number of more/less compliant subjects in a group. The corresponding number of groups is presented in parentheses. An individual is categorized as a more (less) compliant subject if her declared income in each period (out of three periods) is equal to or greater than (less than) the threshold. A group is called a more (less) compliant subjects dominated group if the group consists of at least two more (less) compliant subjects. Whereas 1,000 Lab-points implies full compliance, the median in the first part of treatment I-G-I was 300-Lab-points.

The influence of individual risk preferences on group behavior

Second, we analyze whether the group consists of more or less risk-loving group members. We use the risk measure of the Holt and Laury (2002) task to categorize each participant. In particular, an individual is categorized as a more (less) risk-loving subject if her risk measure is above (below or equal to) the median risk measure (the median in our study is 4 out of 10). A group is called a more (less) risk-loving subjects dominated group if the group consists of at least two (less than two) risk-loving subjects. As a result, we find no significant difference

between the declared income of more risk-loving subjects dominated groups (mean: 185) compared to less risk-loving subjects dominated groups (mean: 133) (MWU-test, two-tailed, p-value > 0.1).

Result 4: *Income declared by a group depends on individual compliance preferences of the group members (measured before group interaction), but does not depend on individual risk preferences (measured before group interaction).*

Female vs. Male Dominated Groups

A robust finding in the tax compliance literature is that women are more compliant than men.[39] At the individual level, we find support for this result in our linear regressions (see tables 3 and 4). When further analyzing the group decisions, we observe that the higher compliance observed at the individual level is indeed able to increase the income finally declared by a group.[40] In particular, the average income reported by the group for female dominated groups (i.e., at least two women in the group) amounts to 243 whereas the mean value is only 56 for male dominated groups (i.e., at least two men in the group). This difference is statistically significant (MWU-test, two-tailed, p = 0.005).

4.6 Arguments in the Group Chat

In this section we analyze the communication in the group chats. Two research assistants independently coded all chats using a predefined codebook containing all variables of interest.[41] In case of an unequal evaluation by the two coders (which was only the case for 9.5% of all messages), a third research assistant independently coded the concerned message. For each variable of interest, the median value across all three coders determined the coding finally used in our analysis. In total, we have 47 groups engaged in 285 chats.[42]

In line with the approach of Kocher et al. (2017), we first categorize the arguments into arguments that are used to encourage compliance and arguments that are used to encourage non-compliance. As our compliance context combines the honesty dimension with the risk dimension, honesty as well as risk can be used as an argument to encourage either compliance

[39] See for example Hasseldine (1999), Kastlunger et al. (2010), and Fochmann and Wolf (2015).

[40] Since evading taxes implies taking risks, our result is in line with Nieboer (2015) who observe that risk aversion in groups is increased with the number of female group members.

[41] The codebook with the variables of interest is presented in the appendix A2.

[42] Groups in treatment I-G-I have 3 separate group chats (3 periods in part 2) and groups in treatment G-G-G have 9 separate group chats (3 parts á 3 periods). Each chat lasts 5 minutes.

or non-compliance. This extends the analysis of Kocher et al. (2017) who focus on honesty and dishonesty arguments.

An argument for non-compliance was mentioned (at least once) by all 47 groups, whereas an argument for compliance was only mentioned by 23 groups (49%). Thus 51% of all groups never mentioned any argument for compliance in their chats. Table 6 shows the number of chats in which arguments for compliance and non-compliance were used. In 16% of all chats (46 of 285), arguments for compliance were mentioned, whereas arguments for non-compliance were used in 52% (149 of 285) of the chats. Similar to the finding of Kocher et al. (2017), we observe that arguments for non-compliance are made significantly more frequently than arguments for compliance (χ^2-test, $p < 0.001$).

Table 6: Number of Chats in which Groups Refer to Compliance and Non-Compliance Arguments

arguments for **compliance** mentioned	arguments for **non-compliance** mentioned		
	yes	no	∑
yes	36	10	46
no	113	126	239
∑	149	136	285

Note: This table shows the number of chats in which arguments for compliance and non-compliance were used.

Furthermore, we analyze which arguments are used to encourage compliance or non-compliance. The majority of arguments refer to risk, money, honesty and taxes. We refer to risk if the message of a group member mentions risk as an argument to encourage compliance or non-compliance (e.g., “I favor to be risky and to declare 0”). Money refers to arguments resting on the monetary consequences of the compliance decision (e.g., “If we declare 0 income, we receive the highest payoff.”). We refer to honesty if honesty is mentioned as a norm or value in order to promote a specific behavior (e.g., “Honesty is the best policy.”). Taxes refers to arguments related to taxes or tax collecting to encourage compliance or non-compliance behavior (e.g., “I think taxes should be paid.”). Figure 8 displays the share of chats, in which these arguments are made in favor of compliance or non-compliance. The argument most frequently used is risk, for both encouraging compliance and non-compliance.

Result 5: *Arguments for non-compliance are made significantly more frequently than arguments for compliance. Arguments referring to risk are the most frequently used arguments to encourage non-compliance and compliance.*

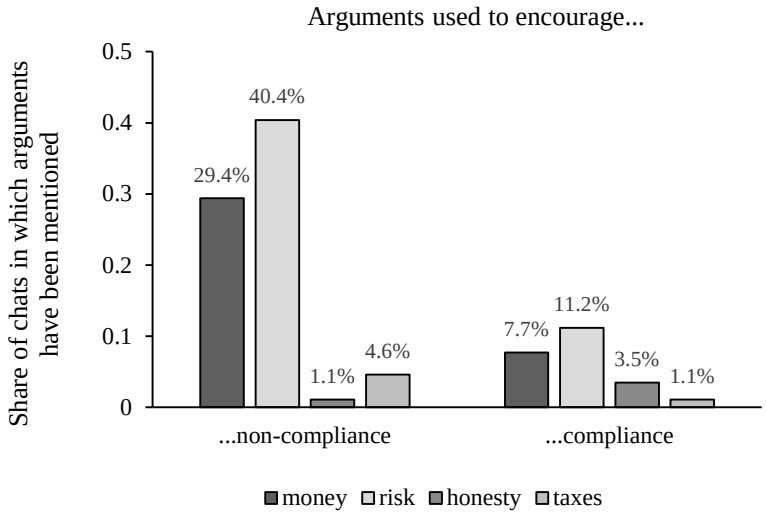


Figure 8: Arguments Used in Group Chats

Table 7 displays our linear regression results with income finally declared by the group in a period as dependent variable. As independent variables, we use dummy variables indicating whether an argument is mentioned in a chat.[43] Whereas model 9 and 10 consider the general use of arguments in favor of compliance or non-compliance, model 11 and 12 distinguish between the different arguments of risk, money, honesty and taxes to encourage either compliance or non-compliance. Model 10 and 12 further control for differences between treatments I-G-I and G-G-G by using a treatment dummy variable which is 1 for treatment G-G-G (0 otherwise).

Models 9 and 10 show that the use of arguments in favor of compliance significantly increases the group's compliance level, whereas the use of non-compliance arguments significantly reduces the compliance level (p-values < 0.01 in all cases). We further find, that the magnitude of the regression coefficient for compliance is about three times as high as for non-compliance. The regression coefficients differ significantly from each other (Wald-test, p-value < 0.001, two-tailed). Thus, arguments for compliance have a much greater impact on the declared income by a group than arguments for non-compliance.

[43] Please notice that our regression results remain robust if we use the frequency of each argument (i.e., how often an argument is mentioned in a chat) as independent variable.

Result 6: *Arguments used to encourage compliance significantly increase group's compliance, whereas arguments used to encourage non-compliance significantly decrease group's compliance.*

Regressing on the different arguments separately (model 11 and 12), we find that the only argument for non-compliance that has a significant influence on the declared income by a group is risk. Neither money, nor honesty and taxes as arguments for non-compliance have a significant influence. When it comes to arguments used to encourage compliance, risk and money significantly increase the declared income. Again, honesty and taxes as arguments do not significantly influence the declared income. Thus, we can summarize that the influence of compliance and non-compliance arguments is mainly driven by the risk argument. When it comes to compliance, money is a further substantial argument.

Result 7: *The influence of communication on the group's compliance behavior is mainly driven by arguments relating to risk. When encouraging compliance, also arguments relating to money have a significant influence.*

Table 7: Group Chats: Linear Regressions with Random Effects
(dependent variable: income declared by group)

	Treatment I-G-I and G-G-G			
Dummy variables (arguments used)	model 9	model 10	model 11	model 12
Compliance	216.34***	217.90***		
	(37.37)	(37.39)		
Non-Compliance	-72.89***	-70.94***		
	(27.26)	(27.31)		
Money (for compliance)			122.00**	120.33**
			(57.80)	(57.61)
Money (for non-compliance)			-37.95	-37.21
			(30.75)	(30.65)
Risk (for compliance)			204.05***	205.83***
			(49.73)	(49.59)
Risk (for non-compliance)			-89.48***	-87.38***
			(29.27)	(29.20)
Honesty (for compliance)			99.94	96.75
			(77.33)	(77.09)
Honesty (for non-compliance)			46.57	40.54
			(123.59)	(123.23)
Taxes (for compliance)			12.87	15.47
			(129.93)	(129.51)
Taxes (for non-compliance)			-45.76	-43.85
			(61.62)	(61.42)
Treatment G-G-G		111.64		108.12
		(71.77)		(70.05)
constant	219.75***	157.81***	230.49***	170.27***
	(39.31)	(55.62)	(37.26)	(53.96)
no. of observations	285	285	285	285
no. of independent groups (clusters)	47	47	47	47
R-squared:				
within	0.098	0.098	0.141	0.141
between	0.378	0.294	0.348	0.283
overall	0.224	0.223	0.219	0.223

Note: In this table, the results of linear regressions are presented with income declared by the group finally in a period as dependent variable (regression coefficients, standard errors in parentheses). Since groups face repeated decisions, we run linear regression models with random effects, where the period is the time variable and the group's identity number is the cross-sectional variable. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

4.7 Summary and Conclusion

Decisions in organizations are often made in groups and unethical behavior is usually associated with the risk of negative consequences for the organization and the involved managers. However, the literature on decision making of groups and individuals does either focus on dishonest behavior without the risk of negative monetary consequences or on risk-taking behavior in situations without ethical concerns. Our contribution is that we extend the

literature by studying the honesty and risk dimension in one setting. In particular, we analyze differences between groups and individuals where unethical behavior can be disclosed and penalized.

We robustly find that groups are less compliant than single individuals. This in line with studies on dishonest behavior observing a dishonesty shift and with studies on risk-taking observing a risky shift in groups. Our results suggest that group communication increases the risk tolerance of the group members and consequently the willingness to enter in risky non-compliant behavior. This provides new evidence that risk is a substantial driver for differences in compliance behavior between groups and individuals. Thus, risk concerns are non-negligible when individuals make compliance decision in groups. The change in norm perception results from the frequent exchange of risk arguments encouraging non-compliance. Whereas individual compliance preferences (measured before group interaction) have an influence on group behavior, individual risk preferences (elicited before group interaction) have not. As we observe a strong decrease in compliance from the individual to the group setting, the group communication with the strong focus on risk arguments seems to increase risk tolerance.

Furthermore, we find that the compliance level is significantly lower in the individual setting after group interaction than in the individual setting before group interaction. Therefore, group interaction induces a negative spill-over effect on subsequent individual compliance. The change in norm perception through communication is still observable after group interaction.

However, we also observe that compliance is significantly higher in the individual setting after group interaction than in the group setting. This finding suggests that the shift in norm perception is not the only driver for the difference in behavior between groups and individuals. Otherwise we would have observed the same compliance level in the group setting and in the subsequent individual setting. Interestingly, the increase of compliance in the individual setting after group interaction was not observable in a context focusing on the honesty dimension only. This finding suggest that our observed effect of group membership on behavior is stronger than the honesty literature without risk concerns predicts. This emphasizes the relevance of the risk dimension in group behavior. To be more precise, Kocher et al. (2017) find a dishonesty shift that is the result of a norm perception shift in groups. They also show that subjects stay on the low honesty level in a subsequent individual setting. This finding suggests that only the shift in norm perception drives the result in the honesty context. In our setup, we observe a norm perception shift in groups as well, but furthermore we find an increase of the compliance level in the subsequent individual setting. Therefore, our results suggest that the change in norm perception is not the sole driver of the results. As social responsibility plays nearly no role in

our setting (as outlined in section 2.3), our results therefore suggest that the tendency that groups are closer to risk-neutrality than individuals also influence decision making in groups.

A categorization of the subjects in our treatment I-G-I reveals that nearly all individuals can be assigned to three types of decision makers. Approx. 40% reveal a constant behavior across all three parts (C-Type). Approx. 20% decrease compliance in the group phase, but increase compliance in the subsequent individual phase again (V-Type) and approx. 30% decrease compliance in the group phase and remain on this low level (L-Type). We therefore contribute to the literature by showing that individuals react differently to group interaction and therefore are heterogeneously in their behavior. But we also show that differences are rather systematic and that a useful categorization is achievable. Furthermore, we find gender differences. Whereas male subjects are mainly C-Types, the V-Type category mainly consists of female subjects. This also implies that the negative spill-over effect of group interaction on subsequent individual compliance is less pronounced for women than for men.

The analysis of the group chat protocols reveals that arguments for non-compliance and for compliance influence compliance behavior in groups. However, non-compliance arguments are made significantly more frequently. Furthermore, we find that the influence of arguments on group behavior is mainly driven by risk arguments. Money is a further substantial argument when compliance is encouraged. Other arguments are not found to significantly influence group behavior. These results extend the findings of Kocher et al. (2017), who show that arguments encouraging ethical behavior (similar to our compliance arguments) reduce lying behavior, but do not find a significant influence of arguments encouraging unethical behavior (similar to our non-compliance arguments). This can be explained by our result that the influence of arguments for non-compliance on behavior is only driven by the risk argument, which does not play a role in the study of Kocher et al. (2017).

References

Allingham, M. G., and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics,* 1, 323–338.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance,* 19, 54–77.

Alm, J., Sanchez, I., and De Juan, A. (1995). Economic and noneconomic factors in tax compliance. *Kyklos,* 48, 3–18.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of economic literature,* 36, 818–860.

Bäker, A., and Mechtel, M. (2015). Peer settings induce cheating on task performance. IAAEU Discussion Paper Series in Economics.

Bär, M., Kempf, A., and Ruenzi, S. (2011). Is a Team Different from the Sum of its Parts? Evidence from Mutual Fund Managers. *Review of Finance,* 15, 359–396.

Baker, R. J., Laury, S. K., and Williams, A. W. (2008). Comparing Small-Group and Individual Behavior in Lottery-Choice Experiments. *Southern Economic Journal,* 75, 367–382.

Bénabou, R., and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review,* 96, 1652–1678.

Blinder, A. S., and Morgan, J. (2005). Are Two Heads Better than One? Monetary Policy by Committee. *Journal of Money, Credit and Banking,* 37, 789–811.

Bliss, R. T., Potter, M. E., and Schwarz, C. (2008). Performance Characteristics of Individually-Managed versus Team-Managed Mutual Funds. *Journal of Portfolio Management,* 34, 110–119.

Bolton, G. E., Ockenfels, A., and Stauf, J. (2015). Social responsibility promotes conservative risk behavior. *European Economic Review,* 74, 109–127.

Bornstein, G., Kugler, T., and Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game: Are groups more "rational" players? *Journal of Experimental Social Psychology,* 40, 599–605.

Bornstein, G., Yaniv, I. (1998). Individual and Group Behavior in the Ultimatum Game: Are Groups More "Rational" Players? *Experimental Economics,* 1, 101–108.

Charness, G., and Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization,* 69, 241–247.

Charness, G., Karni, E., and Levin, D. (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty,* 35, 129–148.

Cheung, S. L., and Palan, S. (2012). Two heads are less bubbly than one: team decision-making in an experimental asset market. *Experimental Economics,* 15, 373–397.

Chytilová, J., and Korbel, V. (2014). Individual and group cheating behavior: A field experiment with adolescents. IES Working Paper.

Cialdini, R. B., and Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology,* 55, 591–621.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology,* 34, 1–7.

Cooper, D. J., and Kagel, J. H. (2005). Are Two Heads Better Than One? Team versus Individual Play in Signaling Games. *American Economic Review*, 95, 477–509.

Cooper, D. J., and Rege, M. (2011). Misery loves company: Social regret and social interaction effects in choices under risk and uncertainty. *Games and Economic Behavior*, 73, 91–110.

Dufwenberg, M., and Dufwenberg, M. A. (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory*, 175, 248–264.

Erat, S., and Gneezy, U. (2012). White lies. *Management Science*, 58, 723–733.

Ertac, S., and Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, 83, 24–30.

Fahr, R., and Irlenbusch, B. (2011). Who follows the crowd – Groups or individuals? *Journal of Economic Behaviour & Organization*, 80, 200–209.

Falk, A., and Tirole, J. (2016). Narratives, Imperatives and Moral Reasoning.

Feri, F., Irlenbusch, B., and Sutter, M. (2010). Efficiency Gains from Team-based Coordination – Large-Scale Experimental Evidence, *American Economic Review*, 100, 1892–1912.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547.

Fochmann, M., and Wolf, N. (2015). Framing and Salience Effects in Tax Evasion Decisions – An Experiment on Underreporting and Overdeducting. SSRN Working Paper.

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior. *Psychological Science*, 20, 393–398.

Gino, F., Ayal, S., and Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization*, 93, 285–292.

Greiner, B. (2004). The online recruitment system ORSEE 2.0 – A guide for the organization of experiments in economics. University of Cologne, Working paper series in economics 10, 63–104.

Harrison, G. W., Lau, M. I., Rutström, E. E., and Tarazona-Gómez, M. (2013). Preferences over social risk. *Oxford Economic Papers*, 65, 25–46.

Hasseldine, J. (1999). Gender Differences in Tax Compliance. *Asia-Pacific Journal of Taxation*, 3, 73–89.

He, H., Martinsson, P., and Sutter, M. (2012). Group decision making under risk: An experiment with student couples. *Economic Letters*, 117, 691–693.

Hofmann, E., Hoelzl, E., and Kirchler, E. (2008). Preconditions of Voluntary Tax Compliance: Knowledge and Evaluation of Taxation, Norms, Fairness, and Motivation to Cooperate. *Zeitschrift für Psychologie/Journal of Psychology*, 216, 209–217.

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

Isenberg, D. J. (1986). Group Polarization: A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology*, 50, 1141–1151.

Janis, I. L. (1972). Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. Oxford, England: Houghton Mifflin.

Kastlunger, B., Dressler, S. G., Kirchler, E., Mittone, L., and Voracek, M. (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D: 4D). *Journal of economic psychology*, 31(4), 542–552.

Kastlunger, B., Kirchler, E., Mittone, L., and Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology*, 30, 405–418.

Kocher, M. G., Pogrebna G., and Sutter, M. (2013). Other-regarding preferences and management styles. *Journal of Economic Behaviour & Organization*, 88, 109–132.

Kocher, M. G., Schudy, S., and Spantig, L. (2017). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science*, 64, 3971–4470.

Kocher, M. G., and Sutter, M. (2005). The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games. *The Economic Journal*, 115, 200–223.

Lahno, A. M., and Serra-Garcia, M. (2015). Peer effects in risk taking: Envy or conformity? *Journal of Risk and Uncertainty*, 50, 73–95.

Maciejovsky, B., Kirchler, E., and Schwarzenberger, H. (2007). Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, 28, 678–691.

Masclet, D., Colombier, N., Denant-Boemont, L., and Lohéac, Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization*, 70, 470–484.

Mazar, N., and Aggarwal, P. (2011). Greasing the palm: Can collectivism promote bribery? *Psychological Science*, 22(7), 843–848.

Mazar, N., Amir, O. and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *Journal of Socio-Economics*, 35, 813–835.

Moffatt, P. G. (2015). Experimetrics: Econometrics for experimental economics. Macmillan International Higher Education.

Muehlheusser, G., Roider, A., and Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economic Letters*, 128, 25–29.

Nieboer, J. (2015). Group member characteristics and risk taking by consensus. *Journal of Behavioral and Experimental Economics*, 57, 81–88.

Prather, L. J., and Middleton, K. L. (2002). Are N+1 heads better than one? The case of mutual fund managers. *Journal of Economic Behavior & Organization*, 47, 103–120.

Rau, H. A. (2015). The disposition effect in team investment decisions: Experimental evidence. *Journal of Banking & Finance*, 61, 272–282.

Reynolds, D. B., Joseph, J., and Sherwood, R. (2009). Risky Shift Versus Cautious Shift: Determining Differences In Risk Taking Between Private And Public Management Decision-Making. *Journal of Business & Economics Research*, 7, 63–78.

Rockenbach, B., Sadrieh, A., and Mathauschek, B. (2007). Teams take the better risks. *Journal of Economic Behavior & Organization*, 63, 412–422.

Schweitzer, M. E., and Hsee, C. K. (2002). Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information. *Journal of Risk and Uncertainty*, 25, 185–201.

Shupp, R. S., and Williams, A. W. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, 118, 258–283.

Sutter, M. (2005). Are four heads better than two? An experimental beauty-contest game with teams of different size. *Economics Letters*, 88, 41–46.

Sutter, M. (2007). Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. *Economic Letters*, 97, 128–132.

Sutter, M. (2009). Deception through telling the truth? Experimental evidence from individuals and teams. *Economic Journal*, 119, 47–60.

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

Weisel, O., and Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 10651–10656.

Whyte, G. (1993). Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes*, 54, 430–455.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115, 157–168.

Appendix

A1 Instructions

Appendix A1 includes the translated instructions (from German). All participants received the general instructions in print. Before the actual experiment was executed, subjects participated in the Holt and Laury (2002) task. The instructions for this task (first experiment) were displayed on the computer screen. After that, participants received the specific instructions for each part of the actual (second) experiment in print.

A1.1 General Instructions

Thank you for participating in this experimental study. For your participation, you will receive a participation fee of 4 Euros.

The experimental study consists of 2 experiments in which you have the opportunity to earn money. Before each experiment, you will receive instructions describing each experiment. Then the experiment starts. After completing the second experiment, you will receive a payout (in addition to the participation fee) which depends on the results of both experiments.

The analysis of the experiment will be anonymous. We will never link your name with the data generated in the experiment. You will not learn the identity of any other participant, neither before nor after the experiment. Also the other participants will not learn your identity. At the end of the experiment, you have to sign a receipt to confirm the payments you received. This receipt will only be used for accounting purposes.

Before we start, we would like to draw your attention to a few important points.

- Please note that you are neither allowed to communicate with other participants nor allowed to leave your desk during both experiments. Please do not look at what other participants are doing.
- Please turn off your mobile phone and store it in your bag.
- Please read the instructions thoroughly.
- It is important that you understand the instructions. Therefore, please do not be afraid to ask questions. If you have any questions, please raise your hand. We will then come to you to answer your questions. Please do not ask questions aloud.
- You can write and make markings on the instructions.
- The calculator and the pen that are lying in front of you can be used.
- Please do not take the instructions home, but return them to us at the end of the study.

Before the first experiment starts, we ask you to fill in a short questionnaire on your computer. After that the instructions for the first experiment will be displayed on your monitor.

A1.2 Instructions for the Holt and Laury (2002) Task

Please choose one of the two lotteries A or B in each of the following 10 decision situations.

You will make a decision for all 10 situations, but your payout from the first experiment is determined only by the one situation that is randomly drawn by the computer after the second experiment.

In each situation, you can either earn 2.00 € or 1.60 € from lottery A and either 3.85 € or 0.10 € from lottery B. The probabilities of winning, however, vary from situation to situation. The further down you move in the table, the higher is the probability of the higher payment and the lower is the probability of the lower payment.

After the first experiment and the second experiment are completed, the computer randomly draws (with the same probability) one of the 10 decision situations. After that, the computer determines your payout from the lottery that you have chosen in this decision situation by a second random draw. For that, the computer uses the probabilities for the higher payment and the lower payment according to the chosen decision situation.

decision	Lottery A	Your decision		Lottery B
		A	B	
1.	2.00 € with 10% or 1.60 € with 90%	o	o	3.85 € with 10% or 0.10 € with 90%
2.	2.00 € with 20% or 1.60 € with 80%	o	o	3.85 € with 20% or 0.10 € with 80%
3.	2.00 € with 30% or 1.60 € with 70%	o	o	3.85 € with 30% or 0.10 € with 70%
4.	2.00 € with 40% or 1.60 € with 60%	o	o	3.85 € with 40% or 0.10 € with 60%
5.	2.00 € with 50% or 1.60 € with 50%	o	o	3.85 € with 50% or 0.10 € with 50%
6.	2.00 € with 60% or 1.60 € with 40%	o	o	3.85 € with 60% or 0.10 € with 40%
7.	2.00 € with 70% or 1.60 € with 30%	o	o	3.85 € with 70% or 0.10 € with 30%
8.	2.00 € with 80% or 1.60 € with 20%	o	o	3.85 € with 80% or 0.10 € with 20%
9.	2.00 € with 90% or 1.60 € with 10%	o	o	3.85 € with 90% or 0.10 € with 10%
10.	2.00 € with 100% or 1.60 € with 0%	o	o	3.85 € with 100% or 0.10 € with 0%

A1.3 Instructions for Main Experiment

A1.3.1 Instructions for Treatment III

A1.3.1.1 Instructions for Part 1

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Corporate employee and corporate income

Imagine you are the employee of a company. Your task is to file the tax return for the company.

As an employee, you receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25 %. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income you want to declare (in the amount of 1000 lab-points). All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \text{ x } \textit{declared} \text{ corporate income}$$

The declaration of the corporate income is the only decision that you make in a single period. In the next period, the decision about the declared corporate income is made again.

Audit of tax return and corporate success

With a probability of 30 %, the provided information on the corporate income is audited. With the counter-probability of 70 %, the information is not audited. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, the company must pay a fine equal to the amount of the unpaid tax.

tax repayment = unpaid tax

Fine = unpaid tax

The unpaid tax is:

$$\text{unpaid tax} = 0{,}25 \text{ x } \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

When deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting company's success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions at your computer. Answering these questions is only a check of your understanding and is not payout relevant.

A1.3.1.2 Instructions for Part 2

The second part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. The second part of the experiment again consists of 3 periods in which you make one decision each.

Corporate employees and corporate income

No changes to the first part of the experiment.

Continue to imagine you are an employee of a company. Your task is to file the tax return for the company.

As in the first part of the experiment, you as an employee receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

No changes to the first part of the experiment.

Therefore, in each period, a tax is again imposed at a rate of 25 %.

The amount of tax to be paid by the company continues to be based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income (which is 1000 lab-points) you want to declare. All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \times \textit{declared}\ \text{corporate income}$$

The declaration of the corporate income is the only decision that you make in a single period. In the next period, the decision about the declared corporate income is made again.

Audit of tax declaration and corporate success

No changes to the first part of the experiment.

It therefore continues to apply that the provided information on the corporate income is audited with a probability of 30 %. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, as in the first part of the experiment, the company must pay a fine equal to the unpaid tax.

Therefore, the company's success continues to result in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

No changes to the first part of the experiment.

Your personal payout in a period continues to consist of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

No changes to the first part of the experiment.

It therefore continues to apply that when deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

A1.3.1.3 Instructions for Part 3

The third part of the experiment is identical to the first and second part of the experiment. This means that you make the same decisions as in the first and second part. The third part of the experiment again consists of 3 periods in which you make one decision each.

A1.3.2 Instructions for Treatment IGI

A1.3.2.1 Instructions for Part 1

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Corporate employee and corporate income

Imagine you are the employee of a company. Your task is to file the tax return for the company.

As an employee, you receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25 %. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income you want to declare (in the amount of 1000 lab-points). All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \times \textit{declared}\ \text{corporate income}$$

The declaration of the corporate income is the only decision that you make in a single period. In the next period, the decision about the declared corporate income is made again.

Audit of tax return and corporate success

With a probability of 30 %, the provided information on the corporate income is audited. With the counter-probability of 70 %, the information is not audited. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, the company must pay a fine equal to the amount of the unpaid tax.

tax repayment = unpaid tax

Fine = unpaid tax

The unpaid tax is:

$$\text{unpaid tax} = 0{,}25 \times \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company’s success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

When deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting company's success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions at your computer. Answering these questions is only a check of your understanding and is not payout relevant.

A1.3.2.2 Instructions for Part 2

The second part of the experiment is identical to the first part of the experiment. The only exception is that you now make your decisions in a triad. Your remuneration therefore also depends on the decisions of other participants.

The second part of the experiment again consists of 3 periods in which you make one decision each.

Group

Together with 2 other, randomly selected participants, you form a triad that stays together during the second part of the experiment. Each of these 3 group members makes the same decisions.

Corporate employees and corporate income

Imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

As in the first part of the experiment, each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company's success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

Therefore, in each period, a tax is again imposed at a rate of 25 %.

The amount of tax to be paid by the company is based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (which is 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by size (from small to large). This also means that if two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \times \textit{declared}\ \text{corporate income}$$

The declaration of the corporate income is the only decision that you and the two other members of your group make in a single period. In the next period, the decision about the declared corporate income is made again.

Before each member submits his binding proposal, the three group members can communicate in writing for a maximum of 5 minutes in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

No changes to the first part of the experiment.

It therefore continues to apply that the provided information on the corporate income is audited with a probability of 30 %. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, as in the first part of the experiment, the company must pay a fine equal to the unpaid tax.

Therefore, the company's success continues to result in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

No changes to the first part of the experiment.

Your personal payout in a period continues to consist of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

While chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

Information about the chat

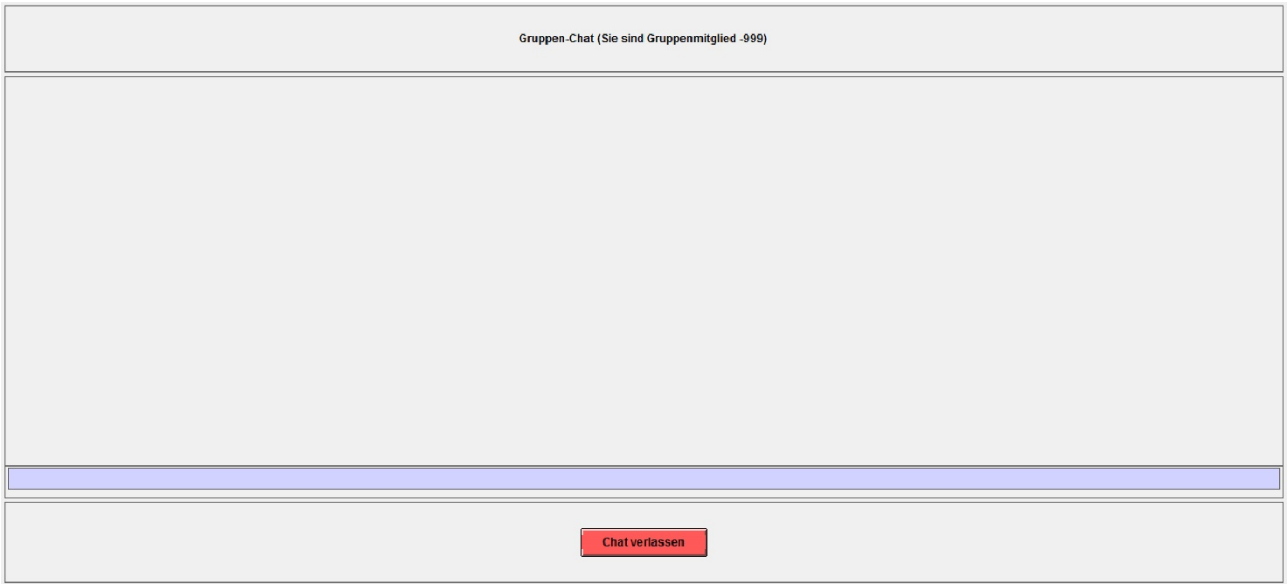
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

You have 5 minutes to exchange information. The group discussion ends after 5 minutes or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do not want to leave the chat, you can click on the button "Back". After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the “enter” key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the “enter” key, that is, when your message appears in the gray box.

A1.3.2.3 Instructions for Part 3

The third part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. Please note, therefore, that you make the decisions **on your own** and not in a group anymore. The third part of the experiment again consists of 3 periods in which you make one decision each.

A1.3.3 Instructions for Treatment GGG

A1.3.3.1 Instructions for Part 1

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions, the decisions of other participants, and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Group

Together with 2 other, randomly selected participants, you form a triad that stays together during the first part of the experiment. Each of these 3 group members makes the same decisions.

Corporate employees and corporate income

Imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

Each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25 %. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (in the amount of 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by size (from small to large). This also means that if

two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \times \textit{declared}\ \text{corporate income}$$

The declaration of the corporate income is the only decision that you and the two other members of your group make in a single period. In the next period, the decision about the declared corporate income is made again.

Before each member submits his binding proposal, the three group members can communicate in writing for a maximum of 5 minutes in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

With a probability of 30 %, the provided information on the corporate income is audited. With the counter-probability of 70 %, the information is not audited. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, the company must pay a fine equal to the unpaid tax.

tax repayment = unpaid tax

Fine = unpaid tax

The unpaid tax is:

$$\text{unpaid tax} = 0{,}25 \times \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company’s success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

While chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting corporate success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions on your computer. Answering these questions is only a check of your understanding and is not payout relevant.

Information about the chat

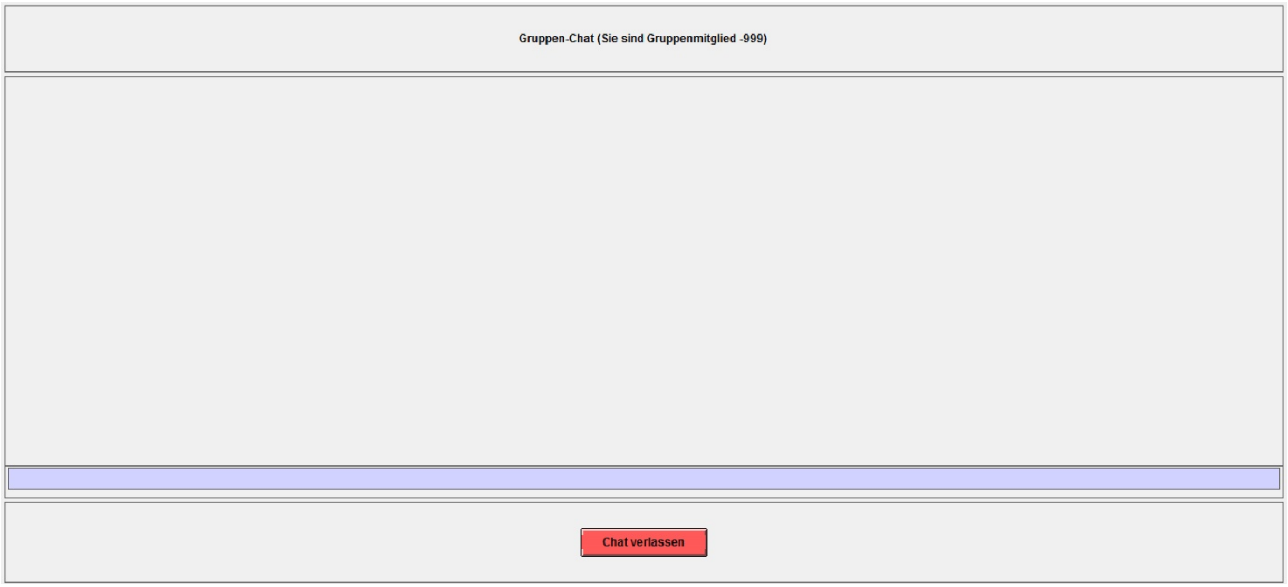
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

You have 5 minutes to exchange information. The group discussion ends after 5 minutes or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do not want to leave the chat, you can click on the button "Back". After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the “enter” key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the “enter” key, that is, when your message appears in the gray box.

A1.3.3.2 Instructions for Part 2

The second part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. The second part of the experiment again consists of 3 periods in which you make one decision each.

Group

Please note that your triad consists of the same group members as in the first part of the experiment and that you therefore interact again in the second part of the experiment with the same participants. Each of the 3 group members makes the same decisions again.

Corporate employees and corporate income

No change to the first part of the experiment.

Continue to imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

As in the first part of the experiment, each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company's success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

No changes to the first part of the experiment.

Therefore, in each period, a tax is again imposed at a rate of 25 %.

The amount of tax to be paid by the company continues to be based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (which is 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by size (from small to large). This also means that if two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25 % of the declared corporate income:

$$\text{tax} = 0{,}25 \times \textit{declared}\ \text{corporate income}$$

The declaration of the corporate income is the only decision that you and the two other members of your group make in a single period. In the next period, the decision about the declared corporate income is made again.

Before each member submits his binding proposal, the three group members can communicate in writing for a maximum of 5 minutes in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

No changes to the first part of the experiment.

It therefore continues to apply that the provided information on the corporate income is audited with a probability of 30 %. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, as in the first part of the experiment, the company must pay a fine equal to the unpaid tax.

Therefore, the company's success continues to result in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

No changes to the first part of the experiment.

Your personal payout in a period continues to consist of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20 % of the company's success. Your personal payout will be as follows:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\ \%\ \text{of the company's success}}_{\text{variable remuneration}}$$

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

No changes to the first part of the experiment.

It therefore continues to apply that while chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

Information about the chat

You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

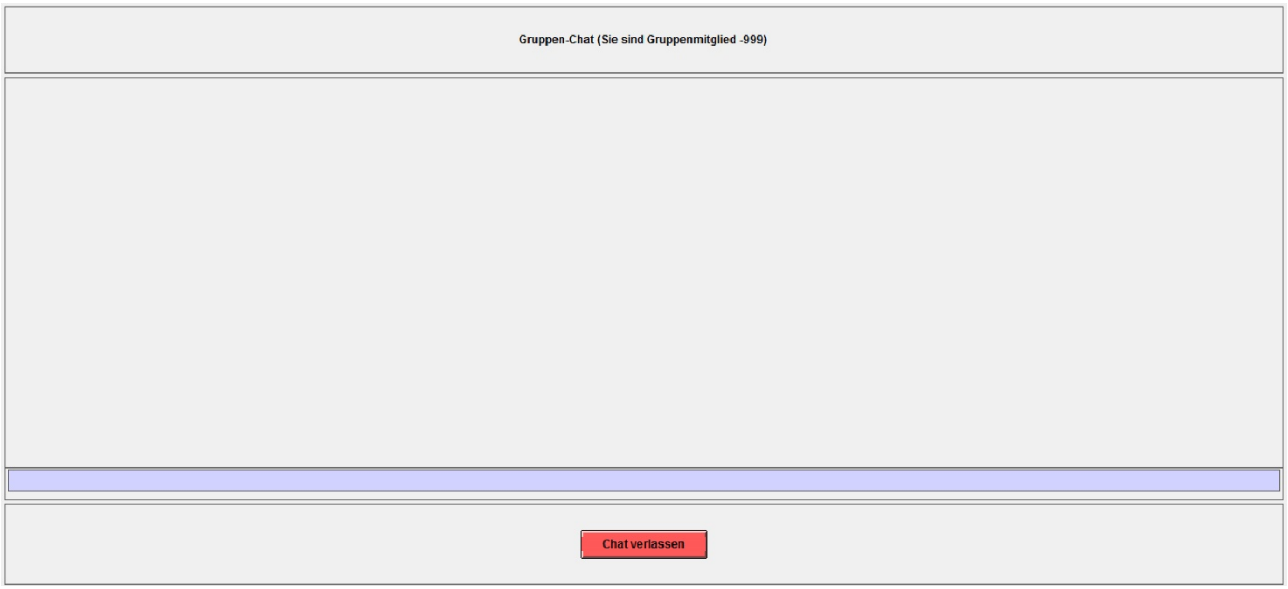
You have 5 minutes to exchange information. The group discussion ends after 5 minutes or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do not want to leave the chat, you can click on the button "Back". After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that

allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the “enter” key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the “enter” key, that is, when your message appears in the gray box.

A1.3.3.3 Instructions for Part 3

The third part of the experiment is identical to the first and second part of the experiment. This means that you make the same decisions as in the first and second part. The third part of the experiment again consists of 3 periods in which you make one decision each.

Please note that your triad consists of the same group members as in the first and second part of the experiment and that you therefore interact again in the third part of the experiment with the same participants.

A2 Codebook

Appendix A2 presents the codebook that was used by the coders.

Arguments used in the group chat:

- **Risk**
  Risk discussed as an argument in the group chat (in general)
  - **Risk_compliance**
    Risk discussed as an argument in favor of compliance
    Example: "I favor to be risky and to declare 0"
  - **Risk_noncompliance**
    Risk discussed as an argument in favor of non-compliance
    Example: "I do not want to take any risks now", "I want to play it safe"

- **Money**
  Money discussed as an argument in the group chat (in general), arguments resting on the monetary consequences of the compliance decision
  - **Money_compliance**
    Money discussed as an argument in favor of compliance
    Example: "I think 170 lab-points are good" (participants receive 170 lab-points when declaring the true income of 1000), "We gain quite a lot if we report honestly"
  - **Money_noncompliance**
    Money discussed as an argument in favor of non-compliance
    Example: "If we declare 0 income, we receive the highest payoff"

- **Honesty**
  Honesty discussed as an argument in the group chat (in general), honesty mentioned as a norm or value
  - **Honesty_compliance**
    Honesty discussed as an argument in favor of compliance
    Example: "In my tax return I'm honest", "Honesty is the best policy"
  - **Honesty_noncompliance**
    Honesty discussed as an argument in favor of non-compliance
    Example: "Let's deceive"

- **Taxes**
  Taxes discussed as an argument in the group chat (in general), arguments related to taxes or tax collecting
  - **Taxes_compliance**
    Taxes discussed as an argument in favor of compliance
    Example: "I think taxes should be paid"
  - **Taxes_noncompliance**
    Taxes discussed as an argument in favor of non-compliance
    Example: "The state does not receive anything"