Experimental Investigations on Tax Compliance and on the Impact of Taxation on Labor Supply

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
– Doctor rerum politicarum –

genehmigte Dissertation
von

M. Sc. Nadja Fochmann
geboren am 28.11.1986 in Berlin

2020

Referenten

Prof. Dr. Kay Blaufus

Prof. Dr. Marina Schröder

Tag der Promotion

24. September 2020

Summary

Diese Dissertation beinhaltet sechs Beiträge über experimentelle Untersuchungen zur Steuerehrlichkeit sowie zum Einfluss von Steuern auf Arbeitsangebotsentscheidungen. Dabei stellen die ersten vier Studien Untersuchungen zur Steuerehrlichkeit bzw. zum Steuerhinterziehungsverhalten dar. Anschließend werden zwei Studien vorgestellt, die den Einfluss von unterschiedlich komplexen Besteuerungssystemen auf das Arbeitsangebot unter realer Arbeitsanstrengung untersuchen.

Die erste Studie untersucht die Auswirkung des Steuergeheimnisses auf die Steuerehrlichkeit in einem Öffentlichen-Gut-Spiel. Es lässt sich ein „Scham-Effekt“ nachweisen, der auftritt, sobald die Mitglieder einer Gruppe für einander sichtbar sind. Dieser erhöht die Steuerehrlichkeit, verschwindet jedoch im Zeitverlauf. Demgegenüber steht der „Ansteckungseffekt“, der die Steuerhinterziehung erhöht, wenn andere Gruppenmitglieder über die Hinterziehung der anderen unterrichtet werden. Dieser hat im gesamten Zeitverlauf bestand, sodass er den „Scham-Effekt“ dominiert. Zudem kann ein allgemeiner Anstieg der Steuerhinterziehung durch den „Ansteckungseffekt“ im Gegensatz zu Entscheidungen unter vollständigem Steuergeheimnis gezeigt werden. Die zweite Studie untersucht die Steuerehrlichkeit von Gruppen im Vergleich zu Individuen. Dabei lässt sich feststellen, dass Gruppen signifikant mehr Steuern hinterziehen als Individuen und dass dieses Verhalten in der Gruppe auch auf spätere Entscheidungen des einzelnen Individuums außerhalb der Gruppe abfärbt. Die dritte Studie untersucht das Steuerhinterziehungsverhalten von Probanden hinsichtlich positiven Einkommens bzw. Gewinnen und negativen Eikommens bzw. Verlusten. Dabei kann gezeigt werden, dass die Steuerehrlichkeit bei der Angabe von positivem Einkommen geringer ist, als bei der Angabe von negativen Einkommen. Dies lässt sich sogar dann feststellen, wenn die Einkommen miteinander verrechnet werden und die Quelle der Steuerhinterziehung dementsprechend bedeutungslos sein sollte. Dieses asymmetrische Entscheidungsverhalten verschwindet erst, wenn die Verrechnung sehr salient dargestellt wird. Die vierte Studie untersucht den Einfluss einer Selbstanzeigemöglichkeit auf die Steuerehrlichkeit. Durch die Einführung dieser Option sind keine Erosion der Steuermoral und damit kein Verdrängungseffekt zu beobachten, da die Steuerehrlichkeit konstant bleibt. Die strategische Nutzung der Option bei ungewissen Prüfungswahrscheinlichkeiten führt jedoch zu einer Zunahme der Steuerhinterziehung, die durch ein geringes Strafgeld deutlich gemildert werden kann.

Die fünfte Studie analysiert die Frage, inwieweit die Ausgestaltung bzw. Komplexität von Steuertarifsystemen einen Einfluss auf die Arbeits-Freizeit-Entscheidung hat. Dabei lässt sich zeigen, dass mit steigender Komplexität des Steuertarifsystems das Arbeitsangebot sinkt. Die sechste Studie untersucht, ob der Zeitpunkt der Pensionsbesteuerung (vor- versus nachgelagert) das Arbeitsangebot und Risikoverhalten beeinflusst. Obwohl beide Steuersysteme zur selben Auszahlung nach Steuern führen, empfinden Probanden ihren Lohn unter der nachgelagerten Besteuerung als fairer, sodass ihr Arbeitsangebot signifikant steigt. Zudem lässt sich zeigen, dass die Risikobereitschaft sinkt.

This dissertation contains six papers on experimental studies on tax compliance and on the effects of taxation on labor supply. In the first four studies, investigations on tax compliance and tax evasion behavior are presented. Subsequently, two studies are presented that examine the influence of different taxation systems and tariff systems on labor supply in real-effort experiments.

The first paper examines the effects of tax s privacy on tax compliance using a public good game. We find a shame effect that occurs when the members of a group are visible to each other. This effect increases tax compliance, but disappears over time. In contrast, we also find a contagion effect that increases tax evasion if the group is informed about the other members' tax evasion without the group members being known. Since this effect persists over time, it dominates the shame effect. Furthermore, we observe an increase in tax evasion due to the contagion effect, in contrast to decisions taken in the context of tax privacy. The second study examines the tax compliance of groups compared to individuals. We find that groups evade taxes significantly more than individuals and that this behavior within the groups also affects later decisions of the individual outside the group. The third study examines the tax evasion behavior of individuals in terms of positive income (or gains) and negative income (or losses). We show that tax compliance is lower for positive income than for negative income. We find this result even when both incomes are offset and the source of tax evasion should be accordingly insignificant. This asymmetric tax evasion behavior only disappears if the offsetting is presented in a very salient manner. The fourth study examines the influence of a voluntary self-disclosure option on tax compliance. By introducing this option, no erosion of tax morale and thus no crowding out effect can be observed, since tax compliance remains constant. However, the strategic use of the option in the case of uncertain audit probabilities leads to an increase in tax evasion, which can be significantly mitigated by a small penalty payment.

The fifth study analyses whether the design or complexity of tax tariff systems influences real work-leisure-decisions. We show that an increased complexity of tax tariffs reduces the labor supply. The sixth study examines whether the timing of pension taxation (immediate or deferred taxation) influences labor supply and risk-taking. Although both tax systems lead to the same after-tax payoffs, subjects in the deferred taxation system perceive their wages as fairer. This significantly increases their labor supply. Furthermore, we find that risk-taking decreases under deferred taxation.

Schlagwörter: Experimentelle Wirtschaftsforschung · Verhaltensorientierte Besteuerung · Steuerhinterziehung · Steuerehrlichkeit · Lügen · Steuergeheimnis · Scham-Effekt · Ansteckungseffekt · Gruppenexperimente · Steuerprüfung · Gewinne · Verluste · Prospect-Theorie · Selbstanzeige · Mentale Buchführung · Steuerwahrnehmung · Fairness · nachgelagerte Besteuerung · Arbeitsangebot · Steuertarifkomplexität · Arbeit-Freizeit-Entscheidung

Keywords: Experimental Economics · Behavioral Taxation · Tax Evasion · Tax Compliance · Cheating · Tax Privacy · Shame Effect · Contagion Effect · Group Experiments · Tax Audits · Gains · Losses · Prospect Theory · Self-Disclosure · Mental Accounting · Tax Perception · Fairness · Deferred Taxation · Labor Supply · Tax Rate Complexity · Real-Effort Experiment

Acknowledgements

My most sincere thanks go to my supervisor and mentor Kay Blaufus who persuaded to take my doctorate during my studies and thereby paved my way into science. He has transferred his passion for experiments to me, has always supported and guided me and relied on my expertise. Without his zeal, his stringent and incredibly inspiring way of working as well as his high level of empathy, I would never have been able to submit this dissertation in this form. I thank the committee members Marina Schröder, Stefan Wielenberg and Michelle Peters for their support.

Moreover, I thank my colleagues at the Institute for Business Taxation at the Leibniz University Hannover for their years of support, sociability and the everlasting joy, fun, teamwork and community. Thereby, especial thanks go to Alexander Schwäbe and Michael Milde who have accompanied and supported me the longest during this time, but also to Sandra Sommer, Janine Jarzembski, Franziska Hoffmann, Yauheniya Filipyeva, Ilko Trenn, Ronny Walch, Jakob Reineke and Julian Simon. Without them, working, research and teaching would have not been the same.

I also thank my co-authors Kay Blaufus, Jochen Hundsdoerfer, Martin Fochmann, Martin Kocher, Jonathan Bob, Philipp E. Otto, Nadja Müller, Chrisian Sielaff, Matthias Sünwoldt and Michael Milde for their enriching collaboration, ideas, and inspiration.

A special thanks goes to my family and friends who supported and believed in me and were always interested in my work and progress. Thereby, I would like to thank above all my parents, parents-in-law, siblings and their families as well as my friends Sophia and Robert, Anne, Christin, Louise and Anja.

My deepest thanks goes to my husband, co-author, former colleague and father of our three sons, Martin, for his initial support in z-tree, his contagious passion for experiments and ingenuity for new research projects. His love, support, thoughtfulness and consideration made it possible to finish my thesis this way.

To Jonas, Jakob and Lukas

Contents

1 Introduction 1

1.1 Motivation and Main Findings 1

1.2 Experimental Economics 3

1.3 Contribution 6

2 The Effect of Tax Privacy on Tax Compliance – An Experimental Investigation 9

3 Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups 10

3.1 Introduction 11

3.2 Related Literature and Hypothesis 12

3.2.1 Risk-Taking Behavior of Individual versus Group Decision Makers 13

3.2.2 Lying and Cheating Behavior of Individual versus Group Decision Makers 14

3.2.3 Empirical Hypothesis 15

3.3 Experimental Design and Sample 16

3.3.1 Decision Task and Payoff Functions 16

3.3.2 Individual and Group Setting 16

3.3.3 Experimental Protocol 17

3.3.4 Sample and Data 18

3.4 Results: Treatment Differences 20

3.5 Treatment I-G-I: Types of Decision Makers and Group Composition 26

3.5.1 Types of Decision Makers in Treatment I-G-I 26

3.5.2 Group Composition in Treatment I-G-I 27

3.6 Arguments in the Group Chat 28

3.7 Summary and Conclusion 32

3.8 Appendix A: Experimental Instructions 33

3.9 Appendix B: Questionnaire 55

3.10 Appendix C: Additional Results 57

3.11 Appendix D: Types of Decision Makers in Treatment I-G-I 59

3.12 Appendix E: Codebook 60

4 Framing and Salience Effects in Tax Evasion Decisions – An Experiment on Underreporting and Overdeducting 62
4.1 Introduction 63
4.2 Related Literature and Main Research Hypothesis 65
4.2.1 Related Literature 65
4.2.2 Main Research Hypothesis 68
4.3 Experimental Design and Tax Evasion Measures 69
4.3.1 Experimental Design 69
4.3.2 Tax Evasion Measures 71
4.4 Experiment 1 and 2: Baseline Experiments 72
4.4.1 Experiment Description 72
4.4.2 Sample 74
4.4.3 Results 74
4.5 Experiment 3 76
4.5.1 No-Offsetting Treatment 78
4.5.2 Offsetting-Low-Salience Treatment 78
4.5.3 Offsetting-High-Salience Treatment 80
4.5.4 Sample 82
4.5.5 Results 82
4.6 Robustness Checks 87
4.6.1 Fear of Further Losses 87
4.6.2 Rewording 88
4.6.3 Tax Evasion Differences over Time 89
4.6.4 Bunching 90
4.7 Discussion 90
4.8 Appendix A: Experimental Instructions 92
4.9 Appendix B: Screenshots from the Experiment 101
4.10 Appendix C: Additional Results 107
5 The Effect of Self Disclosure Opportunities on Tax Compliance – An Experimental Investigation 112

5.1 Introduction 113
5.2 Tax Amnesty and Voluntary Disclosure Research 115
5.3 Hypotheses Development, Experimental Design and Protocol, and Variable Measurement 117
5.3.1 Experimental Design and Hypotheses Development 117
5.3.2 Experimental Protocol and Sample Characteristics 122
5.3.3 Variable Measurement 125
5.4 Results 127
5.4.1 Descriptive Statistics 127
5.4.2 Multivariate Analysis 131
5.4.3 Robustness Analysis 135
5.5 Additional Analyses 136
5.5.1 Tax Morale 136
5.5.2 The Use of the Self-Disclosure Option 137
5.5.3 Complexity and Illegalization 139
5.6 Conclusions 141
5.7 Appendix A: Experimental Instructions 143
5.8 Appendix B: Screenshots of the Experiment 146
5.9 Appendix C: Questionnaire 157
5.10 Appendix D: Derivations 160
5.11 Appendix E: Tables and Figures 164
6 Tax Rate Complexity and Labor Supply – A Real-Effort Experiment 181
6.1 Introduction 182
6.2 Literature Review 183
6.3 Hypothesis and Experimental Design 186
6.3.1 Research Question and Hypothesis 186
6.3.2 Experimental Protocol 187
6.3.3 Experimental Design 188
6.4 Results 191
6.4.1 Tax Perception 191

6.4.2 Accomplishment of the Work Task ... 192
6.4.3 The Effect of Tax Complexity on Labor Supply and Work Performance ... 193
6.4.4 Regression Analysis ... 197
6.4.5 Robustness Check: Labor Supply and Productivity over Time ... 199
6.5 Interpretation and Discussion ... 204
6.6 Limitations and Future Research ... 205
6.7 Appendix A: Experimental Instructions ... 206
6.8 Appendix B: Tax System Sheets ... 207
6.9 Appendix C: Questionnaire ... 208
7 Mental Accounting and the Timing of Pension Taxation ... 210
7.1 Introduction ... 211
7.2 Hypothesis Development ... 213
7.2.1 Work Effort ... 213
7.2.2 Risk-Taking ... 216
7.3 Experimental protocol and sample ... 217
7.3.1 Experimental Protocol ... 217
7.3.2 Sample ... 220
7.4 Results ... 220
7.4.1 Work Effort ... 220
7.4.2 Risk-Taking ... 226
7.5 Robustness Checks ... 231
7.5.1 Work Effort over Time ... 231
7.5.2 Structural Equation Modelling including Risk-Taking ... 232
7.5.3 Instructions Check ... 233
7.6 Conclusion ... 234
7.7 Appendix A: Experimental Instructions ... 236
7.8 Appendix B: Screenshots of the Experiment ... 241
7.9 Appendix C: Questionnaire ... 244
8 Concluding Remarks ... 247

List of Tables

Table 1.1: Essay Overview 7
Table 3.1: Overview of Variables 19
Table 3.2: Treatment I-I-I vs. G-G-G: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: REPORTED INCOME) 24
Table 3.3: Treatment I-I-I vs. I-G-I: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: REPOERTED INCOME) 25
Table 3.4: Types of Decision Makers in Treatment I-G-I 27
Table 3.5: Group Composition in Treatment I-G-I (Dependent Variable: MEAN REPORTED INCOME IN PART 2) 28
Table 3.6: Group Chats: Linear Regressions with Random Effects (Dependent Variable: INCOME REPORTED BY GROUP) 31
Table 3.7 (corresponds to Table 3.2): Treatment I-I-I vs. G-G-G: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: DECLARED INCOME) 57
Table 3.8 (corresponds to Table 3.3): Treatment I-I-I vs. I-G-I: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: DECLARED INCOME) 58
Table 4.1: Example 71
Table 4.2: OLS Regressions for our Two Baseline Experiments 1 and 2 (Dependent Variable: RATIO OF TAX EVASION) 76
Table 4.3: Overview of the Offsetting Mechanisms in both Offsetting Treatments 77
Table 4.4: Tax Evasion Behavior in our Experiment 3 83
Table 4.5: OLS Regressions for our Experiment 3 (Dependent Variable: RATIO OF TAX EVASION) 85
Table 4.6: Treatment Comparisons – OLS Regressions for our Experiment 3 87
Table 4.7: Mean Ratio of Tax Evasion Difference by Period 89
Table 4.8: Descriptive Statistics for Individual Characteristics 107
Table 4.9: Treatment Comparisons – OLS Regressions for our Experiment 3 108
Table 4.10: Treatment Comparisons – OLS Regressions for our Experiment 3 (Participant-Averages) 109
Table 4.11: OLS Regressions for our Robustness Experiments (Dependent Variable: RATIO OF TAX EVASION) 110
Table 4.12: OLS Regressions for our Robustness Test “Bunching” (Dependent Variable: RATIO OF TAX EVASION) 111
Table 5.1: Overview of the Experiment's Periods 124
Table 5.2: Tax Evasion Rates per Period 129
Table 5.3: Regression Results for Tax Evasion 133

Table 5.4: Experiment's Decision Trees and (Expected) Payoffs for all Treatment Variations 164
Table 5.5: Share of Non-Evader per Period 165
Table 5.6: Share of Full Evader per Period 166
Table 5.7: Logit-Regression Results for Non-Evader 167
Table 5.8: Logit-Regression Results for Full Evader 169
Table 5.9: Regression Results for Tax Evasion without ROUND 171
Table 5.10: Logit-Regression Results for Non-Evader without ROUND 173
Table 5.11: Logit-Regression Results for Full Evader without ROUND 175
Table 5.12: Regression Results for Tax Evasion with LAST PERIOD AUDIT 177
Table 5.13: Logit-Regression Results for the Self-Disclosure Use after Initial Tax Evasion 179
Table 5.14: Logit-Regression Results for the Self-Disclosure Use after Initial Tax Evasion Testing for the Occurrence of High Audit Probabilities 180
Table 6.1: Overview of Tax Tariff Systems 189
Table 6.2: Descriptive Statistic in Total and Separated by the Tax Systems 190
Table 6.3: Questionnaire Statistics in Total and Separated by the Tax Systems 191
Table 6.4: Overall Labor Supply, Work Performance and Gross Wage 193
Table 6.5: Labor Supply, Work Performance and Gross Wage Separated by the Tax Systems 194
Table 6.6: Pairwise Treatment Comparisons of Labor Supply and Work Performance 196
Table 6.7: OLS Regression (Dependent Variable: WORKING TIME and TRANSFERRED SHEETS) 198
Table 6.8: Adjusted Labor Supply Separated by the Tax Systems 201
Table 6.9: Pairwise Treatment Comparisons of the Adjusted Labor Supply 202
Table 6.10: OLS Regression (Dependent Variable: ADJUSTED WORKING TIME and ADJUSTED TRANSFERRED SHEETS) 203
Table 7.1: Overview of the Taxation System 214
Table 7.2: Overview of Portfolio Decision Lotteries 219
Table 7.3: Overview of WORK EFFORT Categorized by Treatment 221
Table 7.4: The Impact of Deferred Taxation on Fairness Perceptions 222
Table 7.5: The Impact of Perceived Fairness on WORK EFFORT 223
Table 7.6: Mediation Analysis 224
Table 7.7: Mediation Analysis Including Socio-Demographic Variables and Ability 225
Table 7.8: The Impact of Deferred Taxation on Risk-Taking 226
Table 7.9: Random-Effects Logit Regressions (Dependent Variable: HIGH RISK-TAKING$_t$) 228
Table 7.10: Random-Effects Tobit Regressions (Dependent Variable: RISK-TAKING$_t$) 231
Table 7.11 Mediation Analysis Including Socio-Demographic Variables and Ability, Including Risk-Taking Behavior 233

List of Figures

Figure 3.1: Experimental Design 17
Figure 3.2: Reported Income 21
Figure 3.3: Histograms for Treatments I-I-I and G-G-G (Data Pooled over all Parts) 22
Figure 3.4: Histograms for each Part of Treatment I-G-I 22
Figure 3.5: Arguments Used in Group Chats 29
Figure 3.6: Examples of Constant-Constant-Types 59
Figure 3.7: Examples of Decrease-Increase-Types (V-Shape) 59
Figure 3.8: Examples of Decrease-Constant-Types (L-Shape) 59
Figure 4.1: Tax Evasion Behavior in Experiment 3 (Means Presented) 83
Figure 4.2: Exemplary Screenshot for the Decision Stage of the Baseline ($PI_a \neq |NI_a|$) Experiment 101
Figure 4.3: Exemplary Screenshot for the Information Stage of the Baseline ($PI_a \neq |NI_a|$) Experiment 102
Figure 4.4: Exemplary Screenshot for the Decision Stage of the Offsetting-Low-Salience Treatment 103
Figure 4.5: Exemplary Screenshot for the Information Stage of the Offsetting-Low-Salience Treatment 104
Figure 4.6: Exemplary Screenshot for the Decision Stage of the Offsetting-High-Salience Treatment 105
Figure 4.7: Exemplary Screenshot for the Information Stage of the Offsetting-High-Salience Treatment 106
Figure 5.1: Experiment's Overview 118
Figure 5.2: Distribution of Tax Evasion Levels by Treatment 130
Figure 5.3: Tax Evasion Decision Stage in the First Five Periods 146
Figure 5.4: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the First Five Periods 147
Figure 5.5: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the First Five Periods 147
Figure 5.6: Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the First Five Periods if all Taxes are already Declared Truthfully in the First Five Periods 148
Figure 5.7: Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the First Five Periods if all Taxes are already Declared Truthfully in the First Five Periods 148
Figure 5.8: Information Stage in the First Five Periods if the Subject is not Audited 149
Figure 5.9: Information Stage in the First Five Periods if the Subject is in the Self-Disclosure with Penalty Treatment and Uses the Self-Disclosure Opportunity 149

Figure 5.10: Information Stage in the First Five Periods if the Subject is in the Self-Disclosure without Penalty Treatment and Uses the Self-Disclosure Opportunity 150
Figure 5.11: Information Stage in the First Five Periods if the Subject is Audited 150
Figure 5.12: Information on the Introduction of Two Audit Probabilities 151
Figure 5.13: Tax Evasion Decision Stage in the Last Seven Periods 151
Figure 5.14: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the Last Seven Periods 152
Figure 5.15: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the Last Seven Periods 152
Figure 5.16: Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the Last Seven Periods if all Taxes are already Declared Truthfully 153
Figure 5.17: Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the Last Seven Periods if all Taxes are already Declared Truthfully 153
Figure 5.18: Information Stage in the Last Seven Periods if the Subject is in the Self-Disclosure with Penalty Treatment and uses the Self-Disclosure Opportunity 154
Figure 5.19: Information Stage in the Last Seven Periods if the Subject is in the Self-Disclosure without Penalty Treatment and uses the Self-Disclosure Opportunity 154
Figure 5.20: Information Stage in the Last Seven Periods if the Subject is not Audited 155
Figure 5.21: Information Stage in the Last Seven Periods if the Subject is Audited 155
Figure 5.22: Information Stage on Experiment's Payoff 156
Figure 6.1: Presentation of Labor Supply and Work Performance Separated by Tax System 195
Figure 6.2: Gross and Net Productivity over Time 200
Figure 7.1: Theoretical Model of Fairness Mediating the Relationship between Work Effort and the Tax System 216
Figure 7.2: Timeline of Experimental Protocol 220
Figure 7.3: A Structural Equation Model Linking the Tax System Treatment, Perceived Fairness, and Work Effort 224
Figure 7.4: Distribution of RISK-TAKING for each Treatment 229
Figure 7.5: Number of Correctly Transferred Sheets over Time 232
Figure 7.6: The Working Task (z-Tree Screenshot) 241
Figure 7.7: Information on Actual Work Progress (z-Tree Screenshot) 241
Figure 7.8: Entering a New Sheet Number and Information on Work Progress, Actual Wage before and after Tax, and the Tax Rate for the Immediate Taxation Treatment (z-Tree Screenshot) 242
Figure 7.9: Entering a New Sheet Number and Information on Work Progress and Actual Wage for the Deferred Taxation Treatment (z-Tree Screenshot) 242
Figure 7.10: Investment Decision Stage Period 1 for the Immediate Taxation Treatment (z-Tree Screenshot) 243

Figure 7.11: Investment Decision Stage Period 1 for the Deferred Taxation Treatment
(z-Tree Screenshot) 243

Chapter 1

Introduction

1.1 Motivation and Main Findings

Tax evasion is a severe problem for society. It costs governments billions of dollars worldwide each year, and reduces the amount that can be spent on public services and discriminating those taxpayers who honestly pay their duty. Increasing globalization and the resulting opportunities to hide one's money from the tax authorities abroad raises the governments' suspension to tax evasion, but also enhances the joint effort of states to develop global structures and frameworks to increase tax compliance.[1] In order to address the issue effectively, governments need information on drivers of tax evasion and on proven measures to combat this crime.

This is where the thesis at hand begins. It presents four essays on experiments on tax evasion. The first essay studies whether the abolition of tax privacy fosters tax compliance. Some countries use public disclosures or listing tax evaders in order to induce a potential shame which should occur if friends, neighbors or colleagues find your name on the black list. However, besides the positive shame-effect governments run the risks that public disclosure can cause a contagion effect. This contagion effect describes the potential risk that formerly honest taxpayers are negatively inspired by the tax evasion of others, thus also reducing their tax payments. The experiment shows that we find both effects. However, the shame effect decreases in the long run so that the contagion effect increases tax evasion over time. Additionally, it shows that disclosing tax information anonymously (providing information on the tax evasion of others with revealing their identity) decreases tax compliance compared to providing no information on tax evasion behavior at all.

The second essay analyses whether a group (e.g., boards, teams, departments or committees) which discusses and commonly decides on declaring a company's income is more or less compliant than a single person. It can be shown that groups are less compliant than individuals and that participating in a group's decision also decreases tax compliance for individual choices afterwards. Thus, we again find

[1] The OECD recently developed several programs to enhance tax compliance, such as the "Co-operative Tax Compliance – Building Better Tax Control Frameworks" (OECD, 2016), "The Changing Tax Compliance Environment and the Role of Audit" (OECD, 2017) and the "International Compliance Assurance Programme" (OECD, 2019).

a contagion effect and show that board's decisions on tax compliance underlie an increased probability of dishonesty than individual manager's choices.

The third essay deals with the issue of whether taxpayers treat the decision on tax evasion by underreporting positive income differently than by overdeducting negative income if the opportunities for both measures are identical. The experiments show that subjects increasingly underreport income than overdeducting losses even if both incomes (positive and negative) are accumulated and the source of evasion is therefore irrelevant. Only if the accumulation is made very transparent, the difference in the tax evasion behavior vanishes.

The forth essay examines the impact of the self-disclosure opportunity on tax compliance. I do not find any proof for an erosion of tax morale of formerly honest tax payers (crowding out effect) due to the introduction of a self-disclosure opportunity, but reveals that the opportunity is used strategically so that tax evasion increases if audit probabilities are uncertain. However, if the self-disclosure opportunity is accompanied by a rather small penalty payment, the tax evasion increase is limited to a small extent.

Besides tax evasion, labor supply is another component which is important for a country's economic growth and wealth. Only if the working population makes use of its capacities and resources, economy can prosper according to its possibilities. However, labor supply is exposed to many influential variables, such as leisure alternatives, working atmosphere, gross wages and taxation. Again, experiments are used to separate the influence of taxation on labor supply from all other potential determinants.

In addition to the four essays on tax evasion, the thesis presents two experiments on labor supply. Thereby, the fifth essay examines whether tax rate complexity affects labor supply. We run real-effort experiments and offer subjects a real leisure opportunity. The only difference between the treatments was the diverging taxation's complexity. The experiment revealed that an increased complexity of the tax tariff decreases labor supply significantly.

The sixth essay investigates whether the timing of pension taxation (immediate or deferred taxation) influences labor supply and risk-taking. The experiment is designed in such a way that both systems result in the same after-tax payoffs, so that it should not influence labor supply decisions. However, the experiment reveals that subjects in the deferred taxation system perceive their wage as fairer than in the immediate taxation system. This perception increases their labor supply. Moreover, it shows that risk-taking decreases under deferred taxation.

1.2 Experimental Economics

Experimental economics is a relatively young discipline compared to the established research methods such as model theory and empirical analyses. The conduction of economic lab experiments only started in the late 1940s with just a small number of experiments published per year. Since the 1980s the fraction of experimental papers published in economic target journals increased steadily.[2] Only recently, the Sveriges Riksbank Prize in Economic Science in Memory of Alfred Nobel (shortly also referred as The Nobel Memorial Prize in Economic Sciences) was awarded to Richard H. Thaler in 2017 "for his contributions to behavioral economics" (The Nobel Foundation, 2017) and to Abhijit Banerjee, Esther Duflo and Michael Kremer in 2019 "for their experimental approach to alleviating global poverty" (The Nobel Foundation, 2019). Thus, after Daniel Kahneman who received the Nobel Prize "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" (The Nobel Foundation, 2002) and Vernon L. Smith "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms" (The Nobel Foundation, 2002) in 2002, two of the last three Nobel Prizes in Economic Science were awarded to researchers who have provided major contributions to behavioral and/or experimental economic research.

But what can experiments contribute to the economic discourse that theory and empiricism cannot? Theory attempts to depict action alternatives and processes as well as the resulting decisions in a simplified model so that the optimal behavior of a rationally acting decision-maker can be derived. However, the application of these models to decision-makers shows that they are far from acting completely rationally.[3] Thus, enhanced theoretic models which take behavioral economics into account integrate psychological aspects and insights into economic theory in order to improve predictions of decisions-maker's behavior (e.g., Erard and Feinstein, 1994; Fortin et al., 2007; Dulleck et al., 2016; Dwenger et al., 2016; Fochmann et al., 2019; Farhi and Gabaix, 2020). Experiments can investigate theses theoretic approaches by testing for psychological aspects where observational data is missing or hardly available, so that empirical analyses are not satisfactorily possible. Bounded rationality, social preferences, moral constraints, framing and salience effects, and loss aversion are only a few of these psychological aspects that can explain individual behavior that deviate from economic theory predictions, and are analyzed in the thesis at hand.

Despite the non-deniable importance that is reflected in the high honors of the Nobel Prize within the last years, experimental economics is still confronted with recurring criticism of, inter alia, low external

[2] Falk and Heckman (2009) provide a small data evaluation of published experiments based on the publications in the American Economic Review, Econometrica and Quarterly Journal of Economics based on the years 1980 to 2008.

[3] For fundamental research on bounded rationality and irrational behavior, see e.g., Becker (1962), Kahneman and Tversky (1979), Thaler (1980), Thaler and Shefrin (1981), Kahneman et al. (1982), Simon (1986), Camerer (1989), Laibson (1997), Rabin (1998) and Gilovich et al. (2002).

validity.[4] As described above, experiments are designed to test theory that uses assumptions which often underlie a high degree of abstraction from the real economic world. Thus, critics argue that laboratory experiments are far from reality so that the results gained within experiments cannot be used to make well-founded statements and predictions about reality. Even if the criticism of external validity is admittedly not to be discarded without further ado, it must be noted that successfully published experiments are designed in such a way that they do allow conclusions to be drawn about actual problems and decisions in the respective economic context. However, the weaknesses associated with the abstractions in experiments are countered by the resulting advantages of a high internal validity. Internal validity describes the opportunity of experiments to reproduce theory in the lab. Just because of simplifying real-world's context, experiments are able to exactly test theoretical predictions that they were designed for and can therefore contribute to a better understanding of real phenomena.[5] Thus, due to a high internal validity experiments are optimal for testing theoretical predictions. Furthermore, experiments are used to generate data, which is otherwise hardly available in observational data, such as tax evasion rates or pure work effort.[6]

The thesis at hand presents six experiments four of which were conducted at the Leibniz University Hannover which runs the Leibniz Laboratory for Experimental Economic Research at the Faculty of Economics and Management. The laboratory contains eighteen computerized work-stations which are separated by mobile, sound-absorbing partition walls and one experimenter screen which are all connected by an internal network. Students may register online in a subjects' pool and are invited to experiments via e-mail (Bock et al., 2014).[7] When being invited to an experiment, subjects can enroll in their preferred session ensuring that a subject cannot enroll in the same experiment more than once. At the session's date, subjects are distributed to their workstation by drawing a table tennis ball labelled with the corresponding number. In order to minimize the impact of external effects such as noise, weekdays and time effects, we try to run all treatments in parallel in each session. At the beginning of each session general instructions are read out loud in front of every subject in order to ensure a smooth and undisturbed experiment's process. Thereby subjects are e.g., informed that they must not have verbal or eye contact with each other, should only ask questions in private with the experimenter by raising their arms and to turn off their mobiles. Questions can be asked in private during the whole experiment to reduce uncertainties among the participants. After this general introduction, treatment specific instructions are distributed to the participants that describe the experiment, the task and underlying payout mechanism. The instructions also explain the underlying tax system or tax regulation that differ

4 See e.g., Levitt and List (2007). In this context, Harrison and List (2004) argue in favor of field experiments compared to laboratory experiments.

5 Weimann and Brosig-Koch (2019) p. 22.

6 Falk and Heckman (2009) argue that work effort is not easily measurable and that workers are confronted with a mixture of different incentives that might influence their work effort.

7 We use the software hroot which was developed by Bock et al. (2014).

between the treatments and which reflect the main research focus as we intend to analyze whether it influences the dependent variable, i.e., the tax evasion rate or the labor supply in this thesis.

Many experiments either start with a pre-experimental questionnaire, or one or more training periods. Both measures are meant to ensure that subjects really understood the experiment's task so that distortions and measurement errors due to misunderstandings are reduced and reliable results are obtained. In the pre-experimental questionnaire subjects have to answer questions about the instructions regarding the underlying task, tax system and payout mechanism correctly before they can start the actual experiment. Training periods reflect the precise experiment's periods but are not payout-relevant. Thus, they are better suited to familiarize the participant with the following experimental design than the pre-experimental questionnaire. However, the training advantage on the one has to be evaluated against the disadvantage of the experience gathered on the other hand. Especially, if the experiment intends to examine the dependent variable over time, to establish social norms (such as tax morale), or starts with a special first period, training periods can distort the results.[8] Moreover, they can also not ensure that the subject has read the instructions carefully and is aware of the underlying parameters in a way that the pre-experimental questionnaire can do.

All six experiments presented in this thesis contain incentive payments in order to encourage the subjects to indeed behave and decide as they would do outside the laboratory in their actual everyday decisions. Thereby, all decisions in the experimental tasks are payoff relevant. However, as every decision in the laboratory is based on a relatively high stake of money to give it the appropriate weight we pay the subjects only for one period that is determined by a random draw. Thus, every decision of the subject is potentially relevant and mainly contributes to the final payment.[9]

Besides the requirements of paying the subjects a fair remuneration for participating in the experiment (on average at least 10 Euro per hour), we follow guidelines that are designed to comply with legal provisions and to foster the subjects' pool. Thereby, it is, inter alia, prohibited to process or use data without authorization and to always tell the truth in the course of respectful dealings with each other. This means that no deception is allowed in the experiment, so that the subjects can always rely on what is told them beforehand.[10] Although the investigation of deception has its justification in experimental research, this restriction is necessary in order not to endanger the credibility and trust in future experiments and to distort the results.[11]

---

8 Weimann and Brosig-Koch (2019) p. 48 favor training periods as they provide an opportunity to learn as long as the experiment itself is not meant to study learning effects.

9 Weimann and Brosig-Koch (2019) provide a vivid summary of the payment's importance.

10 Weimann and Brosig-Koch (2019) p. 58 raise the problem that deception pushes the probability of the experiment's acceptance in target journals to zero.

11 Jamison et al. (2008) examine the effect of deception by running the same experiments in two different subjects' pools whereby subjects may only be deceived in one of these pools. They find that deception decreases the willingness to participate in a second experiment and influences the results of this second experiment. Replicating this study, Barrera and Simpson (2012) find no effects of deception on the decisions in the second experiment.

As described above we use student participants in our experiments. Harrison and List (2004) argue that students are not representative surrogate as they are younger and more unexperienced regarding certain decisions, such as investments or tax payments, than an average adult, and they are better educated. This argument has been rejected by Ashton and Kramer (1980), Gemünden (1985), Plott (1987) and Alm (2010) particularly because they are confronted with similar underlying decision situations when they leave school.[12] Although employees in all age groups represent a much better cross-section of society, they are more difficult to recruit and have to be paid more for their participation in the experiment, as their opportunity costs are much higher than for students who can use their free time on campus for an experiment.

Summarizing, even today economic experiments face recurring criticism of, intra alia, external validity or student as pseudo-representative subjects.[13] Despite these reservations economic experiments are a suitable and necessary instrument to test theory and identify behavioral aspects that influence the decision making process. In addition, they offer to collect data which is not observable, or reliably measurable in the real world setting.

1.3 Contribution

The dissertation presents six economic experiments which are displayed in Table 1.1. While four essays present experiments on tax compliance, two essays examine the impact of taxation on labor supply. Three of the presented papers are already published: "The Effect of Tax Privacy on Tax Compliance – An Experimental Investigation" is co-authored by Prof. Dr. Kay Blaufus, Dr. Jonathan Bob and Dr. Philipp E. Otto and was published in 2017 in the European Accounting Review. "Framing and salience effects in tax evasion decisions – An experiment on underreporting and overdeducting" is co-authored by Prof. Dr. Martin Fochmann and was published in 2019 in the Journal of Economic Psychology. The essay "Tax rate complexity and labor supply. A real-effort experiment" is co-authored by Dr. Christian Sielaff and was published in 2016 in the Business Administration Review.[14] Two essays are currently submitted to target journals: "Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups" is co-authored by Prof. Dr. Martin Fochmann, Prof. Dr. Martin Kocher and Nadja Müller and is submitted to the Journal of Economic Behavior & Organization. "The Effect of Self Disclosure on Tax Compliance – An Experimental Investigation" is single-authored and submitted to Journal of Economic Psychology. The sixth paper "Mental Accounting and the Timing of Pension Taxation" is originally co-authored by Prof. Dr. Kay Blaufus, Prof. Dr. Jochen Hundsdoerfer and Dr. Matthias Sünwoldt. At present the essay is undergoing a general overhaul in cooperation with Michael Milde.

[12] This argumentation is supported by the works of Remus (1996), Elliot et al. (2007), Liyanarachchi (2007), Depositario et al. (2009) and Alm et al. (2015).

[13] Another point of criticism that is sometimes raised in the Hawthorne effect which describes the possibility that subjects behave differently in the laboratory as they feel observed (Roethlisberger and Dickson, 1939). However, analyzing the quantitative data collected during the original Hawthorne studies in the 1920s and 1930s, Jones (1992) finds no evidence for the existence of the effect.

[14] Please note, that all my published papers were published under my maiden name "Wolf".

Table 1.1: Essay Overview

Chapter	Title	Co-authors	Current publication Status
Experiments on Tax Compliance			
2	The Effect of Tax Privacy on Tax Compliance – An Experimental Investigation	Kay Blaufus Jonathan Bob Philipp E. Otto	European Accounting Review (2017), 561-580.
3	Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups	Martin Fochmann Martin G. Kocher Nadja Müller	Submitted to Journal of Economic Behavior & Organization
4	Framing and salience effects in tax evasion decisions–An experiment on underreporting and overdeducting	Martin Fochmann	Journal of Economic Psychology (2019), 260-277.
5	The Effect of Self Disclosure on Tax Compliance – An Experimental Investigation	–	Submitted to Journal of Economic Psychology
Experiments on Labor Supply			
6	Tax rate complexity and labor supply. A real-effort experiment	Christian Sielaff	Die Betriebswirtschaft/ Business Administration Review (2016), 65-84.
7	Mental Accounting and the Timing of Pension Taxation	Kay Blaufus Jochen Hundsdoerfer Matthias Sünwoldt	undergoing a major overhaul with Michael Milde

As described above, well designed experiments can shed light on behavioral aspects within the decision making process and test theory where other data is not readily available. For this reason, the thesis at hand examines the fields of tax compliance and work effort experimentally. Inherently, tax compliance or rather tax evasion is naturally not perfectly observable, otherwise it would not exist. In empirical papers which use real world data, tax planning is estimated through other parameters such as the effective tax rate. However, since these estimates can be very inaccurate, it is not possible to make behavioral predictions. Therefore, this thesis contains four essays on tax compliance which is analyzed using different experiments. While the first two essays contain group experiments to, inter alia, observe the group's influence on individual choices, the second two essays contain individual choice experiments. Whereas both group experiments find a contagion effect as the knowledge of the other group members' tax evasion leads to a crowding out effect of formerly intrinsic motivated and honest taxpayers, thus decreasing overall tax compliance, the last essay on tax evasion (individual choice experiment) does not find an erosion of tax morale. However, the triggers of the crowding out effect are

different. While the contagion effect causes a crowding out due to the observation of others' dishonest behavior, we do not find a crowding out effect if a self-disclosure option is introduced which might have reduced tax morale if subjects perceive such an introduction as unjust or legalization of the tax evasion by the tax authority. Moreover we find that tools to combat tax evasion have to be carefully evaluated. First, loosening tax privacy may even decrease tax compliance due to the contagion effect which overrides the shame effect (subjects increase tax evasion in the beginning of the experiment in fear of condemnation) in the long run. Second, the increase of tax evasion due to the strategic usage of a self-disclosure opportunity can be limited if it is linked with even a small penalty payment. Thus, the self-disclosure option may be used to offer subjects a way back to honesty while not endangering tax revenues. Additionally, the thesis finds two drivers that enhance tax evasion. First, subjects evade more taxes by underreporting positive income than by overdeducting negative income unless the offsetting mechanism of both incomes is made vary saliently. Second, we find that groups (boards, teams, departments or committees) exhibit lower tax compliance than individuals.

Besides, the thesis presents two experiments on work effort as it is not easily observable or measurable and influenced by a mixture of different incentives and determinants (Falk and Heckman, 2009). Therefore, the thesis at hand examines the effect of different tax systems on work effort in real-effort experiments while reducing the impact of other factors that might influence work effort, such as working atmosphere, recognition of the work, or working hours. On the one hand, we find that increasing tax tariff complexity reduces labor supply. On the other hand, we examine the influence of two different pension taxation systems (immediate or deferred taxation of wages) and find that subjects perceive their wage fairer if it is taxed at a later time (when the returns of the proceeding investment are taxed). This results in a significantly higher labor supply in the deferred taxation treatment.

Chapter 2

The Effect of Tax Privacy on Tax Compliance – An Experimental Investigation

For copyright reasons this chapter is not available in this published version. This paper was published as Blaufus, K., Bob, J., Otto, P. E., & Wolf, N. (2017). The effect of tax privacy on tax compliance–An experimental investigation. European Accounting Review, 26(3), 561-580.
doi: 10.1080/09638180.2016.1258319

Chapter 3

Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups[*]

Abstract

Unethical behavior in organizations is usually associated with the risk of negative consequences for the organization and for the involved managers if being detected. The existing experimental literature in economics has so far mainly focused on the analysis of unethical behavior in environments that involve no fines or similar monetary consequences. In the current paper, we use a tax compliance framework to study (un-)ethical behavior of individuals and small groups. Our results show that groups are clearly less compliant than individuals. The risk of being detected is the most important aspect in the group communication process when deciding on compliance.

Keywords: Dishonesty · lying · compliance · risk-taking · group decisions · communication · norms · experiment

JEL Codes: C91 · C92 · D03 · H26

[*] This chapter is co-authered by Prof. Dr. Martin Fochmann (Freie University Berlin), Prof. Dr. Martin G. Kocher (University of Vienna) and Nadja Müller (University of Cologne).
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.

3.1 Introduction

Consider an organization that decides on the implementation of provisions of labor laws or on the treatment of tax-relevant circumstances. Such decisions often involve a tradeoff between following the provisions tightly, or deviating from the provisions in relevant aspects to increase profits at the risk of getting detected and having to pay fines or face other forms of punishment. Three elements of such decisions in organizations are crucial: compliance with a (moral) norm, risk of detection, and joint decision making in a group or team.

Several recent examples of behaviors and decisions in organizations that produced massive media attention fall under this definition. Some car producers have allegedly pushed (over the) legal limits of measuring emissions for their diesel engines. Sport organizations have not fully ruled out unethical behavior of their officials. There are many pertinent cases of illegal collusive behavior between firms, cases of companies exploiting their dominant market position, cases of widespread tax fraud of companies (e.g., in connection with the so-called Lux Leaks or the Panama Papers), and cases of financial accounting fraud. The most severe cases make it to the public, but there is of course a continuum of norm violations in terms of severity and impact, meaning that norm-violating behavior in organization is a problem on many levels.

In this paper, we analyze the foundations of group decisions when there is a tradeoff between following a moral norm, resulting in earning smaller profits, and violating the norm, leading to higher profits, with a chance that the norm violation is detected and causing punishment. Building on recent work by, among others, Sutter (2009), Conrads et al. (2013), Gino et al. (2013), Chytilova and Korbel (2017), Muehlheusser et al. (2015), Weisel and Shalvi (2015), and Kocher et al. (2018), we implement a laboratory experiment that uses, without loss of generality, a tax compliance context. Our main innovation is the introduction of a detection probability and a penalty in case of non-compliance to a norm compliance setup that has, thus far, been studied mainly without fines or penalties, when comparing individual and group decisions.

More specifically, each decision maker – an individual or a small group – is a member (set of members) of a tax department that is responsible to file a tax declaration for the organization. Declaring less income than actually earned saves taxes and thus potentially increases profits. When non-compliance is disclosed by an audit, the organization has to pay the evaded taxes plus a penalty. In our individual (group) setting, the tax department consists of one member (three members), but we keep the monetary payoff for each member the same in both conditions, given the same choices. Hence, decisions are directly comparable. Group decisions are the most straightforward implementation of an organizational setup, even though, in their simple form, they have to abstract, e.g., from hierarchies within organizations, to keep the design parsimonious. In order to retain as much experimental control as possible we implement group decisions with communication among group members that take place as anonymous real-time chats.

Our main results are as follows. First, confirming existing evidence in setups without fines or penalties for norm violations, we observe that compliance is significantly lower in the group than in the individual setting, i.e. we confirm what has been termed the individual-group dishonesty shift (Kocher et al., 2018).

Second, arguments regarding risk-taking become very focal in the group communication. Not surprisingly, these arguments are most predictive for the outcome of the group communication, i.e. the final choice taken by the group. The focus of the chat on detection is ex-post proof for the relevance of the risk dimension when studying unethical or dishonest behavior.

Third, in line with the importance of the risk dimension, the mechanisms behind the dishonesty shift here is mainly a shift in risk tolerance of group members, in contrast to the shift in norm perception in Kocher et al. (2018). In both studies, however, it seems that common knowledge of attitudes (towards unethical behavior and towards risk) and mutual encouragement in the pursuit of non-compliant behavior are important drivers of group shifts.

Fourth, we find that group interaction induces a spill-over effect on subsequent individual compliance. Part of the shift in compliance behavior seems permanent, when former group members are asked to take a subsequent individual decision. However, we still observe that compliance is significantly higher in the individual setting after group interaction than in the group setting, suggesting that the shift in norm perception is not the only driver for the difference in behavior between groups and individuals, but that the mutual encouragement in the group (not present in the individual setting) in the non-compliance decision matters as well.

The remainder of this paper is structured as follows: In Section 3.2, we discuss the related literature and develop our main hypothesis. The focus will be on the economics literature on group versus individual decision making with regard to risk (e.g., Rockenbach et al., 2007; Masclet et al., 2009; Harrison et al., 2013) and with regard to unethical behavior as well as on the experimental literature on tax evasion (e.g., Torgler, 2002; Hofmann et al., 2008; Alm, 2012, Dulleck et al., 2016). Section 3.3 describes the details of our experimental design. We empirically analyze compliance behavior and treatment differences in Section 3.4. In Section 3.5, 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 3.6. Section 3.7 concludes the paper and draws implications for situations outside the laboratory.

3.2 Related Literature and Hypothesis

The general literature on differences between individuals and small groups as decision makers is enormous. Most of the experimental literature in social psychology and economics focuses on so-called unitary groups, i.e. groups whose members have to come up with a joint decision after some form of deliberation and do not face any internal conflict in terms of monetary payoffs. However, there might be differences in preferences and attitudes. In the following, we discuss selected work that is relevant to

our setup. Recent surveys that cover a broader spectrum are provided by Charness and Sutter (2012) as well as Kugler et al. (2012).

3.2.1 Risk-Taking Behavior of Individual versus Group Decision Makers

In the 1960s, social psychologists started to investigate the decision behavior of groups and individuals and observed a *risky shift* in groups, meaning that unitary groups tend to take more risk than individual decision makers (see Isenberg, 1986, for an early review). More recent papers, however, report no differences (Harrison et al., 2013) or even provide evidence for a *cautious shift*, implying that group decisions are more risk averse than individual decisions (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 in particular domains (Baker et al., 2008; Shupp and Williams, 2008; He et al., 2012). Compared to individuals, groups seem to be more risk averse in lotteries with a low probability of winning the largest payoff (a high risk lottery), but they are less risk averse when this probability is high (a low risk lottery). Studies investigating risky investment decisions outside the laboratory also fail to provide confirmatory 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 risk-exposed than individually-managed funds. When looking at risk-adjusted performance, Bliss et al. (2008) as well as Prather and Middleton (2002) do not provide evidence for any differences between individually-managed and group/team-managed funds. In contrast, Rockenbach et al. (2007) observe that groups accumulate more expected value at lower risk, i.e. they are better at optimizing. 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 not unambiguous, which could explain why studies fail to find a general tendency in terms of risk-taking. First, there is plenty of evidence that *groups take more rational decisions* than individuals in both strategic and non-strategic tasks (e.g., Bornstein and Yaniv, 1998; Bornstein et al., 2004; Sutter, 2005; Cohen et al., 2009; Feri et al., 2010). Reasons, for example, are that groups are better at learning (Kocher and Sutter, 2005; Cooper and Kagel, 2005; Fahr and Irlenbusch, 2011), reducing behavioral biases (Sutter, 2007; Cheung and Palan, 2012), avoiding extreme decisions (Bär et al., 2011), forming statistical assessments (Blinder and Morgan, 2005), allocating risk (Rockenbach et al., 2007), and they are more correct in Bayesian updating (Charness et al., 2007). Related to risk-taking decisions, more rational decision making induces less noise, but not a general shift in risky decisions.

The second argument for risk-taking differences between groups and individuals is that *social responsibility* might lead to more conservative risk-taking. A variety of studies observe that subjects whose risk decisions affect the payoff of others reveal a reduced willingness to take risks (e.g., Charness

and Jackson, 2009; Reynolds et al., 2009; Ertac and Gurdal, 2012; Pahlke et al., 2015). 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 differences in risk-taking. Conformism refers to the phenomenon that individuals change their behavior to match the behavior of others (Janis, 1972; Cialdini and Goldstein, 2004; Bolton et al., 2015). Recent studies observe that individual decisions under risk can be influenced by the risk preferences of peers such as other group members (Cooper and Rege, 2011; Kocher et al., 2013; Lahno and Serra-Garcia, 2015). Related to conformism, *group polarization* refers to the phenomenon that the outcome of group decision making is more extreme than the average initial tendency of the group members (Isenberg, 1986). Obviously, conformism and group polarization can increase or decrease risk-taking in groups compared to individual decision making, depending potentially on whether the average initial tendencies of the group members leaned towards the risky side or the cautious side.

3.2.2 Lying and Cheating Behavior of Individual versus Group Decision Makers

The number of studies examining unethical behavior has recently been growing quickly. Researchers looked at deception, lying, cheating, tax evasion, corruption, promise breaking, etc., and the vast majority of these studies use either laboratory experiments or field experiments, because field data are not easily available. Investigating differences in unethical behavior between individual decision makers and small unitary groups has attracted attention among researchers only very recently. Thus far, the focus has been on settings in which lying and cheating behavior involve no risk of being caught and punished, in contrast to the setting in the current study. When unethical behavior has no consequences, several papers on lying, cheating and deception provide evidence for a *dishonesty shift* in groups, meaning that groups have a stronger inclination to choose unethically than individuals (e.g., Conrads et al., 2013; Korbel, 2017; Bäker and Mechtel, 2015; Weisel and Shalvi, 2015; Kocher et al., 2018). However, not all studies find differences in unethical behavior between groups and individuals (Sutter, 2009; Azar et al., 2013; Muehlheusser et al., 2015), but we are not aware of any paper that provides evidence for an honesty shift from individuals to groups.

Mainly four reasons for the inclination of groups to behave more unethically than individuals have been discussed. First, groups tend to be *more strategical* than individuals, i.e. they figure out payoff-maximizing strategies in challenging environments more easily than individuals. This is a typical 'wisdom of the crowd' argument. Second, group membership implies the possibility of hiding behind other group members, when it comes to decision making. Hence, *observability of individual actions* within a group is potentially reduced compared to an individual decision making situation. As a

consequence, group members might feel less individual responsibility or accountability for their actions (Mazar and Aggarwal, 2011; Conrads et al., 2013).

The third reason is that *communication* within a group can influence the inclination to behave unethically. Communication allows group members to exchange arguments in favor of or against certain actions. The literature suggests that learning about the preferences and attitudes of others might change *norm perception.* Changes in the norm (perception) might be a consequence of conformism, learning (finding arguments), or 'moral' support by other group members. Although a change in norm perception might increase or decrease unethical behavior in groups, the literature provides evidence that there is a tendency towards more unethical behavior after communication, at least as long as norm violations do not have severe consequences (Gino et al., 2009; Korbel, 2017; Kocher et al., 2018). Fourth, recent studies suggest that groups may have a stronger inclination to behave unethically, when *other people benefit* from their dishonest behavior (Schweitzer and Hsee, 2002; Wiltermuth, 2011; Erat and Gneezy, 2012; Gino et al., 2013; Weisel and Shalvi, 2015). When a group member's unethical behavior increases not only her own payoff but also the payoff of other group members (automatically), this might serve as justification or even a motivation for behaving unethically.[1] 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 unethically in groups is that unethical behavior in a group with communication naturally raises *image concerns*. The intention to behave unethically is usually observable by other group members, which might lead to social image (reputational) concerns (Bénabou and Tirole, 2006; Bénabou et al., 2019; Dufwenberg and Dufwenberg, 2018). It should be noted that it seems difficult to sustain a positive self-image in terms of honesty, once the positive social image has been lost (Gino et al., 2009; Bénabou et al., 2019).

3.2.3 Empirical Hypothesis

The innovation of our study is the combination of the honesty dimension and the risk dimension in a context in which individuals or groups have to decide whether to behave honestly or not. The literature on lying and cheating provides rather conclusive evidence for a dishonesty shift in groups. We expect to replicate this finding. Given mixed results on a risky shift for groups, we remain agnostic with regard to the effect of a risk of being detected and punished. However, it is unclear whether the two dimensions – unethical behavior and behavior under risk – are independent or whether they might interact with each other, potentially to a different extent in individuals than in groups. For instance, the fact that group members can support each other 'morally' could re-inforce the dishonesty shift when a risk of detection is present.

[1] Such behavior is reminiscent of so-called 'white lies'.

We thus expect, in line with the literature, groups to be less compliant than individuals.

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

3.3 Experimental Design and Sample

3.3.1 Decision Task and Payoff Functions

In our experiment, each decision maker – an individual or a group – faces the decision of an employee (of a group of employees) who has to declare the income of her (their) company. The actual income of the company, known by the employee(s), is fixed and amounts to 1,000 Lab-points.[2] The decision maker decides on how much of the actual income should be reported, and all integer values from 0 to 1,000 are allowed. The company is requested to pay a corporate tax of 25% on the reported income.[3] With a probability of 30%, the report is audited. If the audit reveals that the reported income is less than the actual income of 1,000, the company has to pay a penalty that is equivalent to twice the evaded tax, i.e. the company has to repay the evaded tax plus a fine which is equal to the tax amount evaded. If the company is not caught misreporting, there are no consequences. The company's after tax profit is…

…if no audit occurs:

$$Company's\ after\ tax\ profit = 1{,}000 - 0.25 \cdot reported\ income \tag{3.1}$$

…if an audit occurs:

$$Company's\ after\ tax\ profit = 1{,}000 - 0.25 \cdot reported\ income - 2 \cdot 0.25 \cdot (1{,}000 - reported\ income) \tag{3.2}$$

The decision maker's payoff is determined by a fixed remuneration of 20 Lab-points and a variable remuneration that amounts to 20% of the company's after tax profit, i.e.:

$$Decision\ maker's\ payoff = 20 + 0.2 \cdot company's\ after\ tax\ profit \tag{3.3}$$

3.3.2 Individual and Group Setting

In the individual setting, each individual provides an independent tax report. In the group setting, three participants are randomly assigned to a group. The task of the group is the same as in the individual setting (i.e. reporting the company's income). Each group member enters the amount individually. The median of the three proposals determines the income reported by the group. Group members only see the median outcome, but not the individual proposals of the others in their group. Before individual decisions are made, group members are allowed to communicate within the other two group members

[2] The conversion rate from Lab-points to euros is fixed and announced at the beginning: 1 Lab-point corresponds to 0.10 Euro.

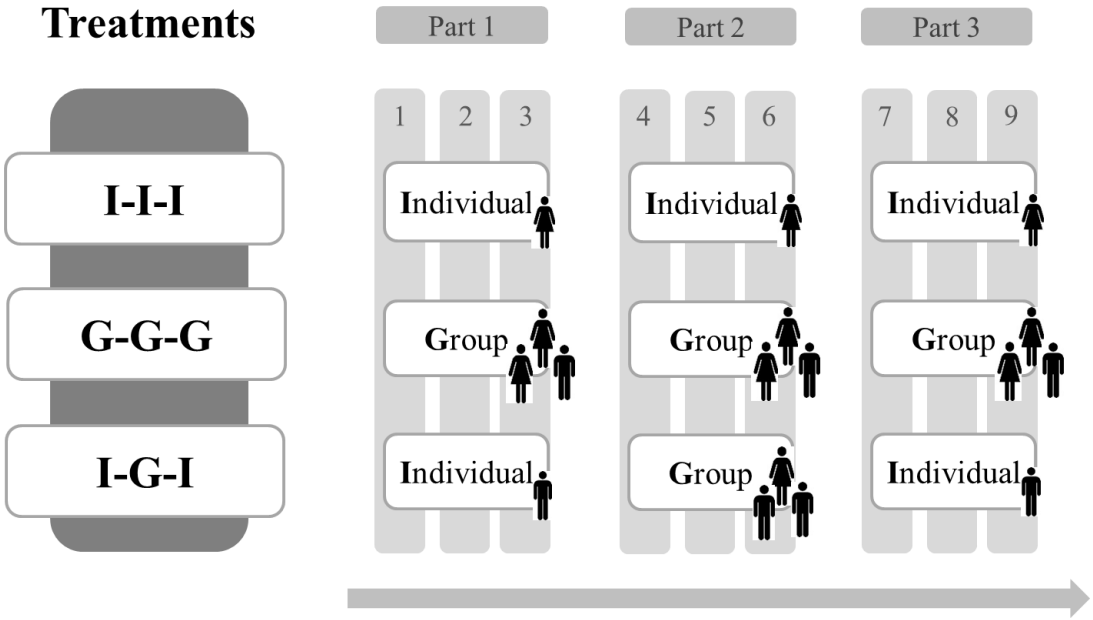
[3] Our chosen levels for the tax rate, audit probability, and penalty are similar to those used in other tax compliance experiments (see, e.g., Alm et al., 1995; Andreoni et al., 1998; Torgler, 2002; Hofmann et al., 2008; Alm, 2012, for excellent literature reviews).

by sending text messages in an anonymous, five-minute chat on the computer screen, without being allowed to reveal their identity, without making side-payments, and without threatening other group members. Each decision maker in the individual setting and in the group setting receives a fixed remuneration of 20 Lab-points and a variable remuneration of 20% of the company's after tax profit; thus, the monetary incentives in the two settings are equivalent.

We implement our settings in three treatments. Each treatment consists of three parts, and each part consists of three consecutive income reporting decisions (i.e. nine decisions in total). Figure 3.1 provides an overview of the treatments. 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. All subjects participate in only one of the three treatments.

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

Figure 3.1: Experimental Design



3.3.3 Experimental Protocol

At the beginning of each part, participants receive written instructions in which all part-related information is presented. The instructions are available in Section 3.8 (Appendix A). 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 will randomly be chosen to determine their individual payoffs.

At the beginning of the experiment, we elicit subjects' willingness to take risks with the Holt and Laury (2002) task (in euro). We use the total number of high risk lottery choices (out of 10) as our proxy for risk attitude. Consequently, subject's willingness to take risk is measured on an 11-point scale, where 0 = not willing to take risk at all, and 10 = strongly willing to take risk. The lottery is resolved at the end of the experiment, and subjects learn the amount that they earned in the lottery after the main part. We obtain further information about individual characteristics of our participants (e.g., gender, age, tax morale, etc.) in a post-experimental questionnaire which is displayed in Section 3.9 Appendix B. At the end of the experiment, each participant receives her total payoff from the experiment plus a show-up fee of 4 euro in cash. In line with most of the tax evasion literature, the instructions are framed in terms of tax decisions. The tax frame should also add to the moral component in decision making.

In the group setting, three subjects are randomly assigned to one group. This assignment was fixed for the rest of the experiment, whenever the group setting is applied. This implies for Treatment G-G-G that a subject is in the same group for the entire experiment. In Treatment I-G-I, a subject stays in the same group for the three decisions of the second part. All messages sent in the chat are received by all group members, and each group member can independently decide to leave the chat. The number of messages sent is not restricted, but the chat automatically ends after five minutes. At the end of each decision situation, in all treatments, each subject is informed about the reported company income, the resulting amount of taxes, whether an audit has been carried out (including a potential penalty), and about the company's after tax profit and the subject's individual profit.

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 her payoff for the outcome with or without an audit. In both the individual and group settings, the "what if" calculator is displayed when subjects decide on the reported income. In the group setting, the calculator is, in addition, also displayed during the chat stage. The experimental software was programmed with z-Tree (Fischbacher, 2007). Participants were recruited with ORSEE (Greiner, 2015).

3.3.4 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, 24.46 euros in approximately 105 minutes (i.e., approximately 14 euros per hour). 48 participants were randomly assigned to Treatment I-I-I, 72 to Treatment G-G-G, and 69 to Treatment I-G-I. Table 3.1 provides an overview of all our variables and presents descriptive results.

Table 3.1: Overview of Variables

Variable	Description	Mean
REPORTED INCOME	income reported 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 LOVING	Holt and 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 MORALE	0 to 9; low tax morale = 0; high tax morale = 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 reported income	0.85

Note: This table presents all variables of our experiment.

3.4 Results: Treatment Differences

Our compliance measure is the income declared by a subject in a given period. Since the actual income was kept constant across periods and treatments (at 1,000), we can use the absolute values of declared income as our variable of interest. We are interested in how the individual willingness to report income truthfully varies across treatments; hence, we use the income each subject declared in the following analyses (if not stated differently). For our non-parametric analyses, we calculated an average per subject in the individual setting (i.e., one independent observation per subject) and an average per group in the group setting (i.e., one independent observation per group).

Figure 3.2 shows averages of DECLARED INCOME for the three treatments. The mean declared income in Treatment I-I-I over all parts is 468 (N=48), whereas it is only 252 (N=24) in G-G-G. The difference is statistically significant (Mann-Whitney U-test; two-tailed; p=0.028). Compliance is significantly lower in the group setting than in the individual setting.

Result 1: *Compliance levels are significantly lower in the group than in the individual setting.*

A similar pattern is observed in each single part. In part 1, mean DECLARED INCOME is 463 (N=48) in Treatment I-I-I, 291 (N=24) in G-G-G, and 392 (N=69) in I-G-I. Whereas the difference between I-I-I and I-G-I is not statistically significant in the first part, the difference between I-I-I and G-G-G is significant at the 10%-level (Mann-Whitney U test; two-tailed; p=0.066), bearing in mind that we use a very conservative test. The difference between I-G-I and G-G-G is not significant (p=0.417).

In part 2, means are 462 (N=48) in I-I-I, 238 (N=24) in G-G-G, and 161 (N=23) in I-G-I. The differences between the individual and group settings are statistically significant (p=0.022 for I-I-I vs. G-G-G and p=0.004 for I-I-I vs. I-G-I). The difference between Treatments I-G-I and G-G-G is not significant (p=0.552).

In part 3, mean DECLARED INCOME is 479 (N=48) in I-I-I, 226 (N=24) in G-G-G and 286 (N=69) in I-G-I. Again, the difference between I-I-I and G-G-G is significant (p=0.014). However, we now also observe a significant difference between I-I-I and I-G-I (p=0.006) and no significant difference between I-G-I and G-G-G (p=0.929).

Treatment I-G-I deserves special attention. Starting out from an average declared income level of 392 in part 1 (the I-part), the level drops to 161 in part 2 (the G-part) (Wilcoxon signed-rank test; two-tailed; p<0.001; N=69). In part 3 of I-G-I (the final I-part), the level increases to 286, but stays significantly below the first I-part.[4] The level is significantly different from part 3 in I-I-I (Mann-Whitney U-test;

[4] The differences between part 2 (161) and part 3 (286) as well as between part 1 (392) and part 3 are statistically significant (p<0.001 for part 2 vs. part 3 and p=0.011 for part 1 vs. part 3; Wilcoxon signed-rank test; two-tailed; N=69).

two-tailed; p=0.006), but not compared to part 3 in G-G-G (p=0.929). Our data provide evidence for a spillover effect in Treatment I-G-I from the G-part to the final I-part.

Result 2: *Group interaction induces a negative spillover effect on subsequent individual compliance: individual compliance is significantly lower after a group interaction.*

Figure 3.2: Reported Income

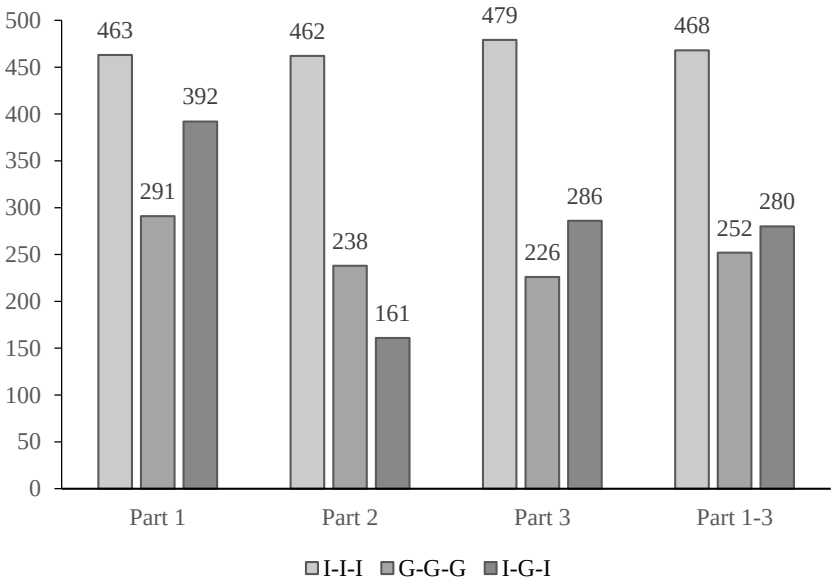


Figure 3.3 shows the histograms for Treatments I-I-I and G-G-G (pooled over all parts), and Figure 3.4 shows the histograms for each part of Treatment I-G-I. As standard in tax compliance experiments, we observe that a relatively high number of subjects chose either to report their income truthfully or to report zero income. Furthermore, we observe spikes for round values (i.e. 100, 200, etc.) in all treatments. Coordination in groups is extremely high, despite the fact that it is not required in our design, since the median proposal is implemented. Nonetheless, almost 90% of the proposals within a group are, on average, the same.

Figure 3.3: Histograms for Treatments I-I-I and G-G-G (Data Pooled over all Parts)

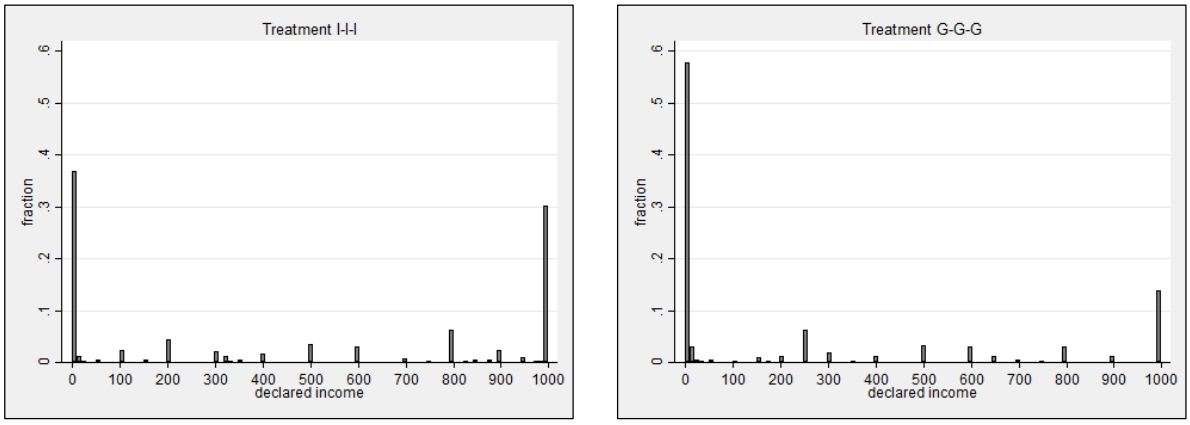
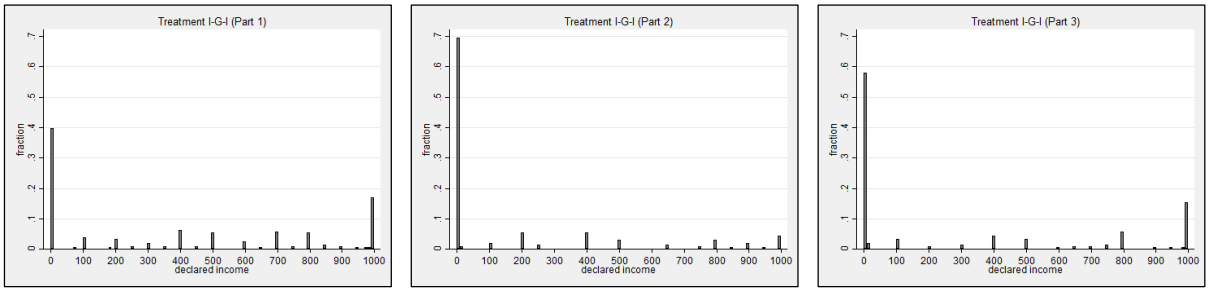


Figure 3.4: Histograms for each Part of Treatment I-G-I



In the following we corroborate our results by running linear regressions that take background variables and the natural correlation structure of data from group interactions into account. We use the DECLARED INCOME by each subject in every period as the dependent variable. As subjects face repeated decisions over several periods, we run multi-level mixed effects linear regressions to capture more than one level of dependence.[5] 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.[6]

Table 3.2 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 our reference, the coefficient of the treatment dummy measures the difference between Treatments I-I-I and G-G-G. We observe a significant lower level of declared income in Treatment G-G-G and therefore confirm our result 1.

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 any additional difference between Treatments I-I-I and G-G-G in parts 2 and 3, respectively.

5 A detailed description of multi-level modelling is, for example, provided in Moffatt (2015).

6 As robustness tests, we rerun all regressions as random-effects panel regressions (panel variable: subject ID, time variable: period) with standard errors clustered on the group level. All results are robust to this variation.

Statistical significance between our two part dummies and our two interaction terms was checked by Wald tests, and the resulting p-values are reported in 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 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 3.2, we show that reported income decreases over the three parts in Treatment G-G-G, whereas it is almost constant 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 models 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 an income declaration had been audited in the previous period, and 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 such as gender, age, etc. We incorporate all 19 individual variables reported in Table 3.1. We show the coefficient of the dummy variable FEMALE in Table 3.2 (which takes the value of 1 if the decision was made by a female, and 0 otherwise). All other individual variables are not displayed.[7] Again, we observe the very similar results as in 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.[8]

[7] The complete set of all regression results are presented in Section 3.10 (Appendix C).

[8] 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 have been audited.

Table 3.2: Treatment I-I-I vs. G-G-G: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: REPORTED 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			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 REPORTED INCOME as dependent variable (regression coefficients, standard errors in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

In Table 3.3, we use the same approach and specifications to analyze the differences between Treatments I-I-I and I-G-I. In models 1 and 3, we observe a significant lower compliance level in Treatment I-G-I. However, models 2 and 4 reveal that the treatment difference is only significant in parts 2 and 3, when one would expect it to be. 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 2 and 4). Again, compliance is lower in the group than in the individual setting (result 1).

In both models 2 and 4, 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 3.2, we find a positive and significant effect of FEMALE and a negative and significant effect of LAST PERIOD AUDIT:

Table 3.3: Treatment I-I-I vs. I-G-I: Multi-Level Mixed Effects Linear Regressions (Dependent Variable: REPOERTED INCOME)

	Treatment I-I-I vs. I-G-I			
	model 1	model 2	model 3	model 4
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			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 REPORTED INCOME as dependent variable (regression coefficients, standard errors in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

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

3.5.1 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), as a group member (part 2), and after group interaction (part 3). We analyze whether we find 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 and analyze the differences across parts. The following patterns are possible, where the first (second) term describes the transition from part 1 to 2 (from part 2 to 3): Decrease-Increase, Decrease-Constant, Decrease-Decrease, Increase-Increase, Increase-Constant, Increase-Decrease, Constant-Increase, Constant-Constant, and Constant-Decrease. We define the transition from part 1 to 2 the following way: If the average declared income of a subject drops (increases) by at least 100 Lab-points from part 1 to 2, then this is classified as a 'decrease' ('increase'), and as 'constant' otherwise. The same logic applies to the transition from part 2 to 3.[9] Examples are shown in Section 3.11 (Appendix D).

The empirical patterns in our data are very straightforward: 88% (61 out of 69 subjects) can be assigned to three types. The relatively most frequent type is the Constant-Constant-Type (26 subjects). 22 out of these 26 subjects can be further classified as low Constant-Constant-Types as their reported income did not exceed 100 in any of the three parts. Almost as many subjects are represented by the Decrease-Constant-Type (20 subjects), following an L-shape. The third type among the most frequent types is Decrease-Increase, following a V-shape, with 15 subjects being classified as such.[10] The entire distribution of subjects in our categorization of different decision maker types is displayed in Table 3.4. Interestingly, we observe substantial gender differences. 51% of all men are categorized as low Constant-Constant-Types, whereas the corresponding figure for women is only 12%. Women are predominantly Decrease-Increase- (32%) and Decrease-Constant-Types (32%). The distribution of the types of men and women is significantly different (chi-squared test; $p<0.05$; $N=61$). Women react to deciding as an individual or as a group member, whereas men do so to a lesser extent. We cannot distinguish whether this is a consequence of women being more responsive to the decision environment or whether this is a floor effect for the in part 1 already less compliant men.

[9] As robustness checks, we also used transition levels of 150 and 200 Lab-points. All results are qualitatively robust to this variation.

[10] In that context Blaufus et al. (2017) also find a contagion effect, i.e, that participation in a group has the effect of reducing tax compliance of formerly honest subjects in the short and long term. In this dissertation the respective paper is presented in Chapter 2.

Result 3: *Almost all subjects can be assigned to three types of decision makers: low Constant-Constant- (32%), Decrease-Constant- (29%), and Decrease-Increase-Types (22%). Male subjects are much more often classified as low Constant-Constant-Types than females; women are more often classified as Decrease-Constant- and Decrease-Increase-Types.*

Table 3.4: Types of Decision Makers in Treatment I-G-I

	All Subjects (69 subjects)				Men Only (35 subjects)				Women Only (34 subjects)			
Types	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
Constant-Constant	26	38%	26	38%	19	54%	19	54%	7	21%	7	21%
High (i.e., > 500)			4	6%			1	3%			3	9%
Low (i.e., <= 100)			22	32%			18	51%			4	12%
Decrease-Increase			15	22%			4	11%			11	32%
Decrease-Constant	37	54%	20	29%	14	40%	9	26%	23	68%	11	32%
Decrease-decrease			2	3%			1	3%			1	3%
Increase-Increase			1	1%			1	3%			0	0%
Increase-Constant	5	7%	3	4%	2	6%	1	3%	3	9%	2	6%
Increase-Decrease			1	1%			0	0%			1	3%
Constant-Increase	1	1%	1	1%	0	0%	0	0%	1	1%	1	1%
Constant-Decrease			0	0%			0	0%			0	0%

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

3.5.2 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, and the sex composition of groups. We use the MEAN REPORTED INCOME IN PART 2 in Treatment I-G-I (group decisions) as the dependent variable and regress it on the group members' MEAN REPORTED INCOME IN PART 1 (individual decisions) in the same treatment, the MEAN NUMBER OF RISKY CHOICES of the group members in the risk elicitation (the Holt and Laury, 2002, task), and the NUMBER OF FEMALES in the group. Table 3.5 presents the regression results. Both the MEAN REPORTED INCOME from PART 1 and the NUMBER OF FEMALES in the group[11] show up significantly in the regressions. The coefficient for risk has the correct sign, but it is far from being significant.

[11] A robust finding in the tax compliance literature is that women are more compliant than men (e.g., Kastlunger et al., 2010; Fochmann and Wolf, 2019). At the individual level, we find support for this result (see our linear regressions in Table 3.2 and Table 3.3). Here, we show that group composition matters, in addition to the difference in compliance on the individual level.

Table 3.5: Group Composition in Treatment I-G-I (Dependent Variable: MEAN REPORTED INCOME IN PART 2)

	Individual Compliance in Part 1	Individual Risk Preferences	Gender
	model 1	model 2	model 3
MEAN REPORTED INCOME IN PART 1 (of the three group members)	0.65*** (0.22)		
MEAN NUMBER OF RISKY CHOICES (of the three group members)		-19.90 (51.08)	
NUMBER OF FEMALES (in group)			109.95** (46.80)
CONSTANT	-102.33 (96.52)	237.82 (222.62)	-8.91 (83.95)
No. of Observations	23	23	23
R-squared	0.299	0.007	0.208

Note: In this table, the results of linear regressions are presented with mean reported income in part 2 as dependent variable (regression coefficients, standard errors in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

3.6 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 pre-defined codebook containing all variables of interest.[12] In case of different evaluations by the two coders (which was only the case for 9.5% of all messages), a third research assistant coded the concerned message independently. For each variable of interest, the median value across all three coders determined the coding that was finally used in our analysis. In total, we have 47 groups engaged in 285 chats.[13]

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 or non-compliance. This extends the analysis of Kocher et al. (2018), who lack the risk dimension.

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%). 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 support a risky choice, i.e. to declare zero"). Money refers to arguments associated with the monetary consequences of the compliance decision (e.g., "If we declare

[12] The complete codebook with all variables of interest can be found in Section 3.12 (Appendix E).

[13] Groups in Treatment I-G-I have three separate group chats (three decision situations in part 2; 3 x 23 groups = 69 chats) and groups in Treatment G-G-G have nine separate group chats (three parts á three decision situations; 9 x 24 groups = 216 chats). Each chat lasts five minutes.

zero 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 (normative) arguments related to taxes or tax collection (e.g., "I think taxes should be paid."). Figure 3.5 displays the share of chats, in which these arguments are brought forward in the chats. The most frequent type of argument used is associated with risk, for both encouraging and discouraging compliance.

Result 4: *Arguments for non-compliance are made significantly more frequently than arguments for compliance. Arguments referring to risk are the most frequent arguments to encourage or discourage compliance.*

Figure 3.5: Arguments Used in Group Chats

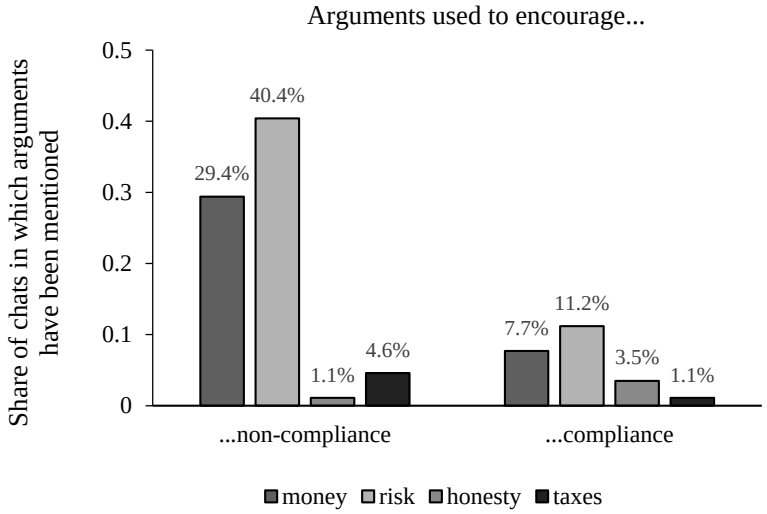


Table 3.6 displays linear regression results with INCOME REPORTED BY GROUP in a period as dependent variable. As independent variables, we use dummy variables indicating whether a specific type of argument is mentioned in a chat.[14] Whereas models 1 and 2 consider the general use of arguments in favor of COMPLIANCE or NON-COMPLIANCE, models 3 and 4 distinguish between the different specific types relating to RISK, MONEY, HONESTY, and TAXES to encourage either COMPLIANCE or NON-COMPLIANCE. Models 2 and 4 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, and 0 otherwise.

Not surprisingly, the use of arguments in favor of compliance significantly increases a group's compliance level, whereas the use of non-compliance arguments significantly reduces the compliance level ($p<0.01$ 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, $p<0.001$, two-tailed). Thus, arguments for compliance have a

[14] Our results remain qualitatively unchanged if we use the frequency of each argument (i.e., how often an argument is mentioned in a chat) as independent variable.

much greater impact on the reported income by a group than arguments for non-compliance, supposedly because compliance in a group is harder to achieve.

> ***Result 5:*** *Arguments in the group chat 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 (models 3 and 4), we find that the only arguments for non-compliance that has a significant influence on the reported group income are arguments related to risk. When it comes to arguments used to encourage compliance, arguments related to risk and money significantly increase the declared group income.

> ***Result 6:*** *The influence of communication on the group's compliance behavior is mainly driven by arguments relating to risk.*

Table 3.6: Group Chats: Linear Regressions with Random Effects (Dependent Variable: INCOME REPORTED BY GROUP)

	Treatment I-G-I and G-G-G			
Dummy variables (arguments used)	model 1	model 2	model 3	model 4
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 INCOME REPORTED BY 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. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

3.7 Summary and Conclusion

The paper analyzes group decisions when there is a tradeoff between following a moral norm, resulting in earning smaller profits, and violating the norm, leading to higher profits. The innovation compared to existing recent work is the introduction of a chance that the norm violation is detected and that norm-violation might be punished. Decisions makers in our experiment are either individuals or groups of three members. They are thought of as a tax department at an organization, responsible for filing a tax declaration. Declaring less income than actually earned saves taxes and thus potentially increases profits. When non-compliance is disclosed by an audit, the organization has to pay the evaded taxes plus a penalty.

We confirm existing evidence from setups without fines or penalties for norm violations: groups declare significantly smaller amounts than individuals, i.e. they behave less honestly. Importantly, the risk dimension is the most important aspect in the decision making procedure within groups. It is most often discussed in the group communications, and it has an effect on the outcome of the declared group income.

We also find conclusive evidence on a spill-over effect of group decision making on subsequent individual compliance. Part of the individual-group dishonesty shift in compliance seems permanent even when former group members are asked to take a subsequent individual decision. However, we still observe that compliance is significantly higher in the final individual setting after the group interaction than in the group setting itself.

First, the replication of previous results in a different frame and with different decision making rules than in previous papers is a comforting outcome. Groups indeed seem inclined to behave more dishonestly than individuals. Second, the hitherto neglected risk domain appears to be relevant. If there is a risk of being detected, it seems that the individual-group dishonesty shift becomes even more important. Future work should look at different detection probabilities to further corroborate this preliminary result.

3.8 Appendix A: Experimental Instructions

Appendix A 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 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.

A2 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%

A3 Instructions for Main Experiment

A3.1 Instructions for Treatment I-I-I

A3.1.1 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:

tax = 0.25 x *declared* 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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable} \text{ remuneration}}$$

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.

A3.1.2 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:

tax = 0.25 x *declared* 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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable}\text{ remuneration}}$$

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).

A3.1.3 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.

A3.2 Instructions for Treatment I-G-I

A3.2.1 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:

tax = 0.25 x *declared* 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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable} \text{ remuneration}}$$

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.

A3.2.2 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:

tax = 0.25 x *declared* 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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable}\text{ remuneration}}$$

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

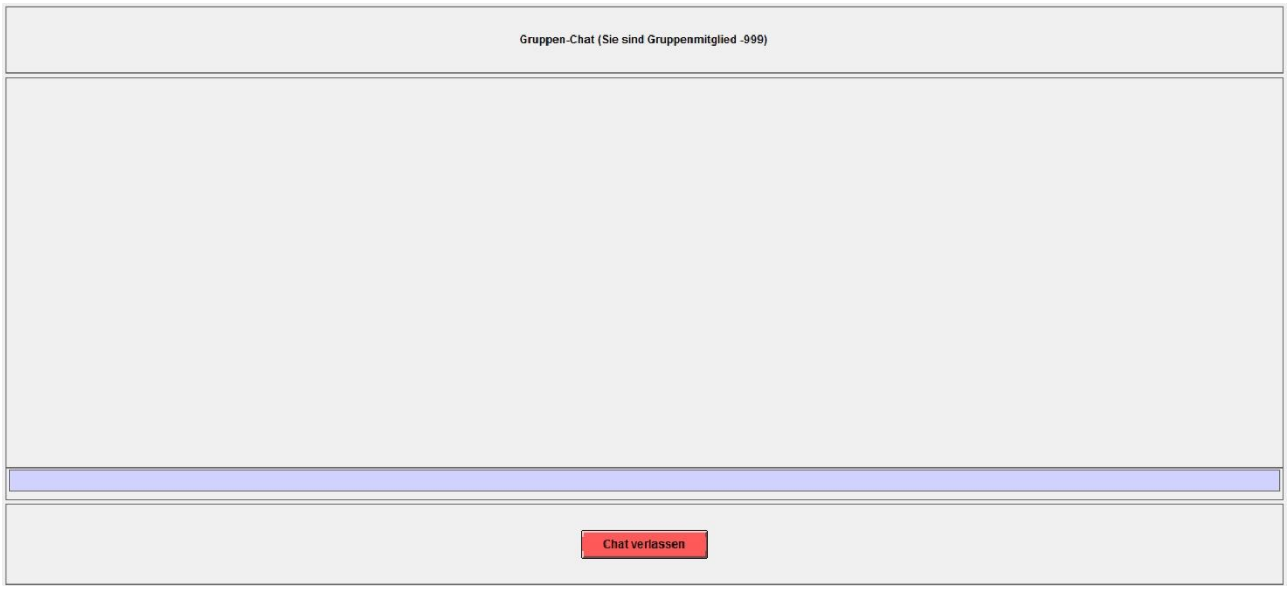
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

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.

A3.2.3 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.

A3.3 Instructions for Treatment G-G-G

A3.3.1 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:

tax = 0.25 x *declared* 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:

$$\text{unpaid tax} = 0.25 \text{ x } \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable}\text{ remuneration}}$$

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

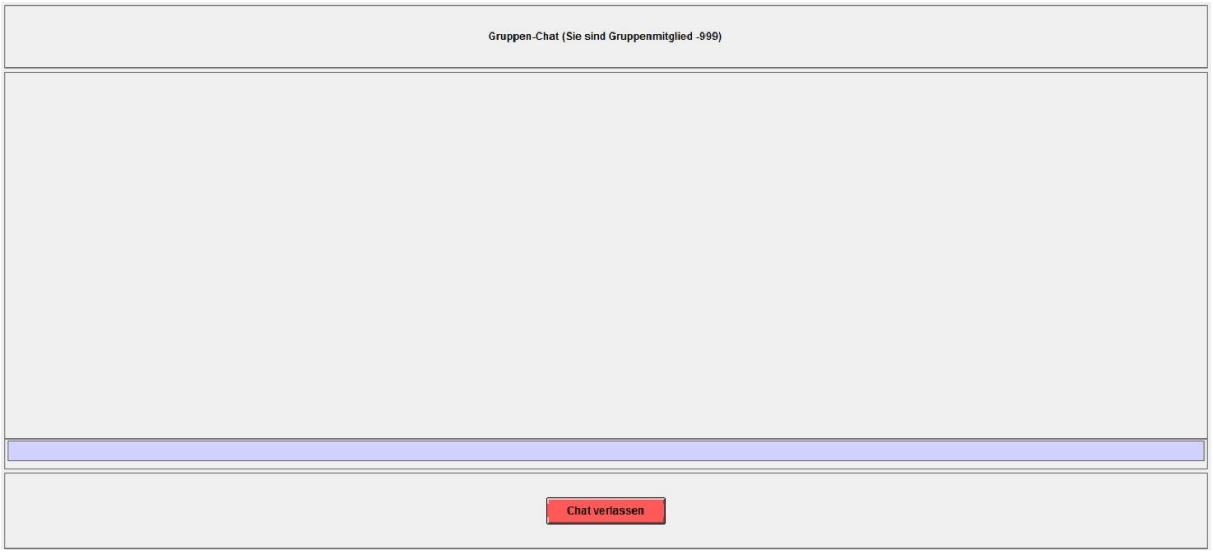
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

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.

A3.3.2 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:

tax = 0.25 x *declared* 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:

$$\text{payout in a period} = \underbrace{20}_{\text{fixed remuneration}} + \underbrace{20\% \text{ of the company's success}}_{\textit{variable}\text{ remuneration}}$$

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

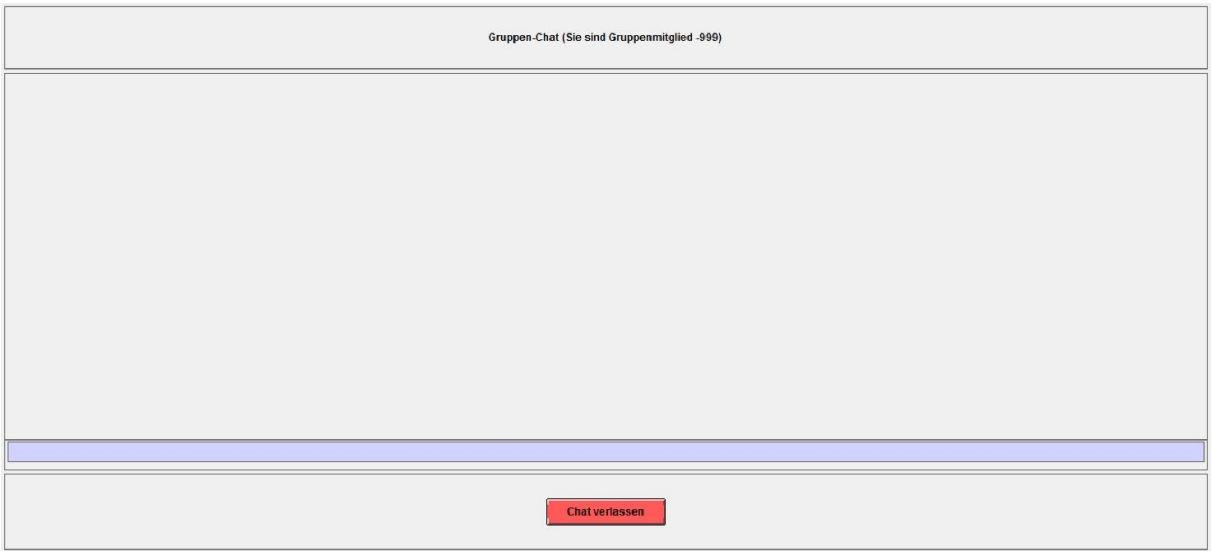
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

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.

A3.3.3 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.

3.9 Appendix B: Questionnaire

How old are you? ☐

Are you female or male?

☐ Female ☐ Male

Have you attended more than one lecture from the Department of Economics?

☐ Yes ☐ No

What qualification are you aiming at right now?

☐ Bachelor
☐ Master
☐ Diploma
☐ state examination
☐ Other ☐

Do you ever have submitted a tax return?

☐ Yes – once
☐ Yes – twice
☐ Yes – more than twice
☐ No
☐ I do not know

Do you have knowledge of tax law (e.g. through apprenticeship or through tax law lectures)?

☐ Yes ☐ No

Please judge the following statements.

	You are not allowed to do this on any terms									It is all right in any case
Evading taxes if you have the possibility to do so.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

	On no account									On any account
If somebody does something good to you, are you content to reciprocate even if it was not agreed upon beforehand?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
If someone treats you badly, do you also treat this person badly?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Very unfair									very fair
How do you rate the contribution and control system which was applied within the experiment?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Not complex at all									Very complex
How complex do you perceive the decision periods (fiscal decisions)?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please state, which emotions you have perceived in the course of the experiment.

	It is not true at all									It is absolutely true
I have perceived joy.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived anger.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived anxiety.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived guilt.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

What is your monthly disposable income (after rent)? ☐

How often do you pray per week?

- ☐ Never
- ☐ 1 – 2 times
- ☐ 3 – 5 times
- ☐ Daily
- ☐ Several times a day

3.10 Appendix C: Additional Results

This section presents the complete set of regression results. Table 3.7 corresponds to Table 3.2 and Table 3.8 to Table 3.3.

Table 3.7 (corresponds to Table 3.2): 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			154.66***	153.82***
			(42.40)	(42.42)
RISK LOVING			0.89	0.68
			(10.44)	(10.45)
AGE			-0.27	-0.27
			(2.40)	(2.40)
ECONOMICS			-52.99	-52.78
			(37.90)	(37.92)
BACHELOR			26.22	26.43
			(35.17)	(35.18)
TAX EXPERIENCE			48.56	48.31
			(35.24)	(35.26)
TAX KNOWLEDGE			38.09	37.77
			(53.27)	(53.30)
TAX MORALITY			6.88	6.88
			(6.68)	(6.68)
POSITIV RECIPROCITY			4.52	4.46
			(9.18)	(9.19)
NEGATIVE RECIPROCITY			-6.32	-6.25
			(5.63)	(5.64)
FAIRNESS			8.56	8.59
			(5.98)	(5.99)
DECISION COMPLEXITY			-9.08	-9.12
			(9.68)	(9.69)
JOY			-4.02	-4.04
			(5.82)	(5.83)
ANGER			-10.03**	-9.93**
			(4.42)	(4.42)
FEAR			11.93	11.79
			(7.71)	(7.71)
GUILT			11.16	11.09
			(9.23)	(9.23)
INCOME			-0.04	-0.04
			(0.08)	(0.08)
RELIGIOUS			15.95	16.15
			(39.78)	(39.80)
WHAT IF CALCULATIONS			0.72	5.07
			(6.60)	(7.15)
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). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 3.8 (corresponds to Table 3.3): 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) WITHIN PART			22.17**	13.75
			(10.27)	(10.41)
FEMALE			178.48***	177.84***
			(49.32)	(49.07)
RISK LOVING			-25.29*	-27.94*
			(15.29)	(15.24)
AGE			9.39***	9.15***
			(3.33)	(3.31)
ECONOMICS			-56.88	-53.76
			(46.97)	(46.74)
BACHELOR			-25.98	-27.44
			(45.46)	(45.24)
TAX EXPERIENCE			43.31	45.80
			(52.26)	(52.00)
TAX KNOWLEDGE			102.13	102.53
			(71.28)	(70.93)
TAX MORALITY			12.30	11.61
			(9.88)	(9.84)
POSITIV RECIPROCITY			-22.02*	-21.79*
			(13.12)	(13.05)
NEGATIVE RECIPROCITY			1.64	0.66
			(10.11)	(10.06)
FAIRNESS			9.08	7.89
			(7.92)	(7.88)
DECISION COMPLEXITY			-9.22	-8.80
			(12.59)	(12.53)
JOY			-4.26	-3.55
			(8.58)	(8.53)
ANGER			7.20	6.74
			(8.16)	(8.12)
FEAR			14.25	14.06
			(10.86)	(10.81)
GUILT			-4.25	-3.12
			(8.46)	(8.43)
INCOME			-0.28***	-0.27***
			(0.09)	(0.09)
RELIGIOUS			-45.51	-42.73
			(55.70)	(55.42)
WHAT IF CALCULATIONS			6.98*	3.12
			(4.12)	(4.21)
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). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

3.11 Appendix D: Types of Decision Makers in Treatment I-G-I

Figure 3.6 to Figure 3.8 display examples for the most frequent types of decision makers observed in our experiment.

Figure 3.6: Examples of Constant-Constant-Types

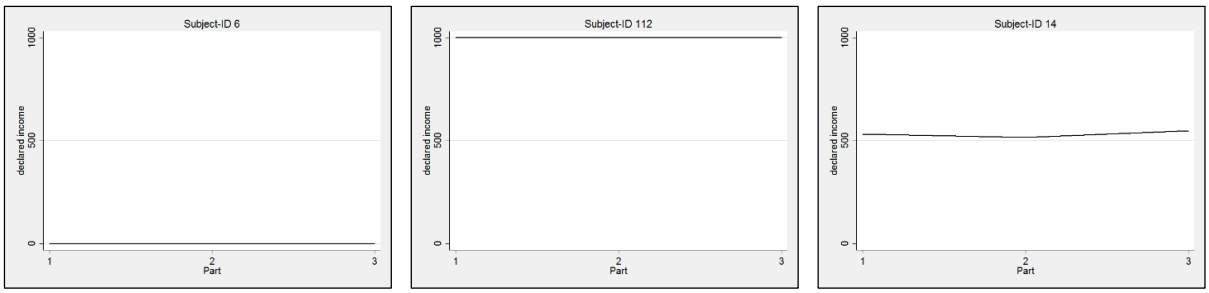


Figure 3.7: Examples of Decrease-Increase-Types (V-Shape)

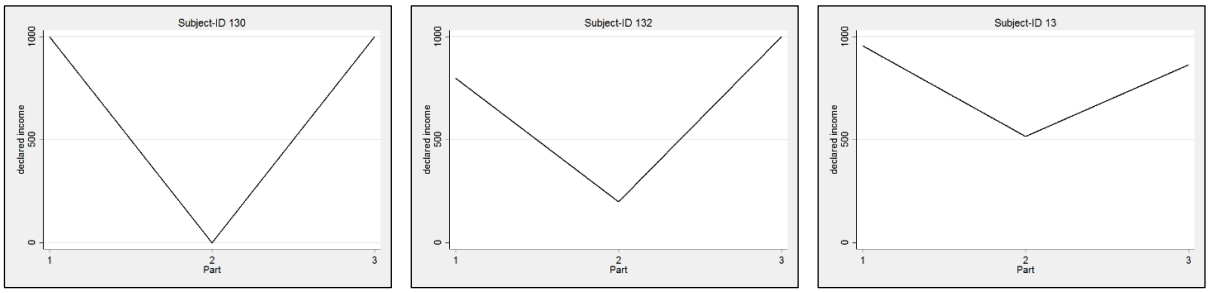
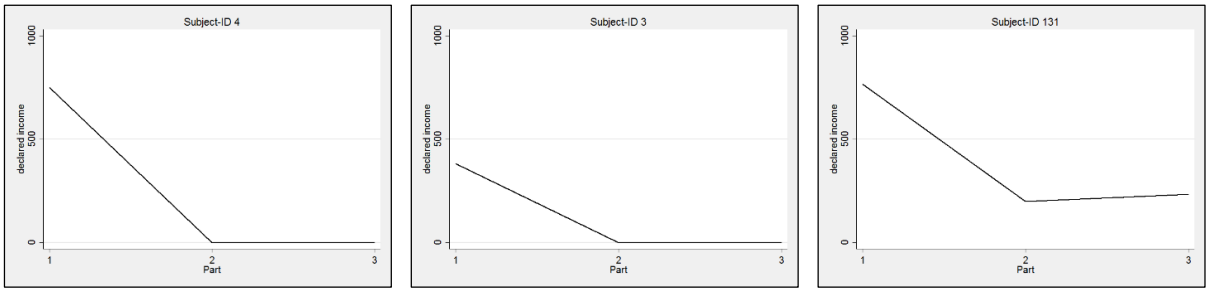


Figure 3.8: Examples of Decrease-Constant-Types (L-Shape)



3.12 Appendix E: Codebook

This section presents the codebook that was used by the coders.

Arguments used:

- **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 do not want to take any risks now", "I want to play it safe"

  - **Risk_noncompliance**

    Risk discussed as an argument in favor of non-compliance

    Example: "I favor to be risky and to declare 0"

- **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: "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"

Chapter 4

Framing and Salience Effects in Tax Evasion Decisions – An Experiment on Underreporting and Overdeducting[*]

Abstract

Using different controlled and incentivized experiments, we analyze whether taxpayers are more willing to evade taxes by underreporting positive income (e.g., business or nonbusiness income) than by overdeducting negative income (e.g., deductions, credits, or losses). We robustly observe an asymmetric tax evasion behavior. Specifically, individuals are less compliant in case of positive income. This result is robust to a variation in which the after-tax payoffs from the cases with positive and negative income are offset against each other – even though offsetting reduces the asymmetric effect significantly. In an experimental environment in which the interaction of positive and negative income reporting is made very saliently and in which we consequently expect that subjects decide on both tax evasion decisions jointly, the asymmetric effect vanishes. We therefore provide evidence that 1) tax evasion behavior is asymmetrically in case of positive and negative income reporting, 2) offsetting reduces this asymmetric effect and 3) the salience of income interaction plays an important role in tax evasion decisions.

Keywords: Tax evasion · cheating · prospect theory · behavioral taxation · experimental economics

JEL Codes: C91 · D14 · H26

[*] This chapter is co-authored by Prof. Dr. Martin Fochmann (Freie University Berlin) and was published in the Journal of Economic Psychology (2019), 260-277.
We especially thank Nadja Müller for helpful research assistance. Moreover, we thank the editor Laetitia Mulder and two anonymous referees as well as Abdolkarim Sadrieh, Kay Blaufus, Ken Klassen, John Barrick, Jochen Hundsdoerfer, Sebastian Eichfelder, Friedrich Heinemann, Philipp Doerrenberg, Ulrich Schreiber, Christoph Spengel, Michael Overesch, Frank Hechtner, Jost Heckemeyer, Maximilian Todtenhaupt, Harald Amberger, Andreas Knabe, Gerhard Schwödiauer, Matthias Braune, Corinna Treisch, André Renz, Terese Z. Nopress, Alexander Schwäbe, and participants of the 2016 Workshop on Current Research in Taxation in Bonn and Mannheim, 2016 MaTax Conference, 2015 EAA Conference, 2015 VHB Conference, 2014 GfeW Conference, 2014 arqus Conference, research seminar at the University of Magdeburg and Free University of Berlin for helpful comments and suggestions.

4.1 Introduction

To ensure that tax evasion is combatted effectively, researchers, institutions, and governments has undertaken certain effort to quantify and analyze tax evasion behavior. In the U.S., for example, the National Research Program (NRP) was implemented in 2001 to analyze 46,000 tax returns by experienced auditors in order to determine audit strategies and to estimate the tax gap. Slemrod (2007) analyzes the individual income tax data in more detail and depicts that taxpayers overdeduct $15 billion offsets to income (4% of the true amount) and overdeduct $17 billion credits (26% of the true amount), while they underreport $56 billion nonbusiness income (4% of the true amount) and $109 billion business income (43% of the true amount). Based on this analysis underreporting income exceeds overreporting of credits and offsets to income for the individual income tax. The question of our paper is: Is there a systematic (preference induced) difference between under- and overreporting? Or more specifically: Are tax evaders more willing to evade taxes by underreporting positive income (e.g., business or nonbusiness income) than by overdeducting negative income (e.g., deductions, credits, or losses)?

We decided to pursue lab experiments to study these research questions. Even though we are aware that this might reduce external validity[1], we have chosen this approach as this allows us to provide a clean test of a systematic asymmetric tax evasion behavior driven by preferences. In reality, under- und overreporting are usually no perfect substitutes for example due to opportunity constraints or different audit and detection risks.[2] As benefits and costs of both evasion strategies may vary across taxpayers (individuals or companies), limited or unlimited entities, industry sectors, countries, and degrees of business internationalization, a clean identification of a systematic—and preference induced—divergence in tax evasion behavior is limited. In our study, however, we keep the benefits and costs of evasion constant for both tax evasion strategies. An identified divergence effect then adds to the monetary effects induced by different benefits and costs.

Although the body of related literature is rather small, some papers—in addition to Slemrod (2007)—offer evidence of asymmetric tax evasion behavior. Most closely connected to our contribution are Cameron et al. (2008), Kern and Chugh (2009), and Grolleau et al. (2016), who show—in a neutral context—that individuals are more willing to cheat in case of a loss frame than in case of a gain frame.

---

1 Alm et al. (2015) report evidence that laboratory experiments in the field of tax compliance exhibit external validity. They show that behavioral pattern of subjects in the laboratory corresponds to the decision making in naturally occurring setting. For a detailed discussion of the use of laboratory experiments for tax research issues see also Alm (2010).

2 For example, self-employed persons and employees have different opportunities for tax evasion. Whereas employees have only a very limited possibility of underreporting income if the employer is forced by law to report the employee's income to the tax authority, self-employed individuals have the ability to either underreport income or overdeduct expenses or deductions. Furthermore, underreporting income is usually related to not recording a transaction whereas overdeduction of expenses is rather related to falsely recording a transaction. This difference can per se imply different benefits and costs of both evasion strategies which prevents both strategies of being perfect substitutes.

If we transfer these results to a tax evasion context, we could hypothesize a higher tax evasion level in a situation where a taxpayer has to pay a tax (loss frame) and a lower level in a situation where a taxpayer receives a tax refund (gain frame). These expectations are supported by the literature on the income tax withholding phenomenon (see, for example, Robben et al., 1990a; Chang and Schultz, 1990; Schepanski and Shearer, 1995; Elffers and Hessing, 1997; Jackson and Hatfield, 2005; Engström et al., 2015), that suggests that taxpayers are likely to evade more taxes when they are confronted with an additional tax payment and less taxes when they are confronted with a tax refund. Although these results look promising, the studies do not analyze whether taxpayers more often evade taxes by underreporting positive income or by overdeducting negative income.

With our study, we contribute to this literature and shed further light on this discussion. We conduct three controlled and incentivized experiments with 205 participants in total. In each decision situation of our experiments, a subject has to make two decisions: one tax evasion decision in case of positive income and one in case of negative income. In the first case, a subject is confronted with a positive amount of income and has to pay a tax in accordance with her reported positive income. In the second case, a subject is confronted with negative income and receives a tax refund in accordance with her reported negative income.

In our framework, we use a rather abstract and general understanding of the terms positive and negative income. For example, positive/negative income can refer to business and nonbusiness income of an individual taxpayer (e.g. employee, sole proprietor, or freelancer) or to income of a corporation or partnership. Furthermore, positive income can represent a positive aggregated amount of different income sources (like the positive amount of business income after expense deduction, the adjusted gross income, or the final tax base) or positive income from a single income source (e.g., revenues or wages). Negative income can stand for deductible expenses and credits or for a negative aggregated amount of different income sources (when for example expenses outweigh the company's revenues). Positive and negative income can also refer to two different companies with one generating a gain whereas the other generates a loss. In case of group reliefs or fiscal unities, both companies are allowed to file a consolidated tax return so that the loss can be offset against the gain. However, if the requirements are not met, an offset is usually not allowed. In case of two different branches or departments within one company, the loss of one branch/department is naturally offset against the gain of the other branch/department when the (consolidated) tax return is completed by the company.

Our main results are threefold: First, we find that individuals evade more in case of positive income than in case of negative income. Second, this result is robust to a variation in which the after-tax payoffs from the cases with positive and negative income are offset against each other (Offsetting-Low-Salience experiment). In this setting, it does not matter whether the individual evades taxes through underreporting or overdeducting as both opportunities are perfect substitutes. A divergent tax evasion behavior should therefore not be observed. However, the experimental design is modeled in such a way

that the decision on positive income is clearly separated from the negative income decision. Consequently, the offsetting of positive and negative income is not presented saliently and the resulting interaction of income reporting—that leads to the arithmetical irrelevance of whether taxes are evaded by underreporting positive or overdeducting negative income—is difficult to identify. We thus hypothesize that individuals still decide on the tax evasion of positive and negative income independently and separate both decisions like in the two baseline experiments. Again, we find a divergent tax evasion behavior as in the baseline conditions. However, we also observe that offsetting reduces the magnitude of the divergence effect significantly.

Third, if the offsetting is presented very saliently and the interaction of income reporting can therefore be identified very easily (Offsetting-High-Salience experiment), we expect subjects to be more aware of the irrelevance of the income channel that is used for tax evasion. Although there is no arithmetical difference compared to the Offsetting-Low-Salience experiment, we hypothesize that subjects decide on both tax evasion decisions jointly in this experiment variation. Consequently, no different tax evasion behavior should be observed for positive and negative income. As hypothesized, we show that our initial effect of an asymmetric tax evasion vanishes. Taken all these results together, we are able to provide evidence that 1) tax evasion behavior is asymmetrically in case of positive and negative income reporting, 2) offsetting reduces this asymmetric effect and 3) the salience of income interaction plays an important role in tax evasion decisions.

4.2 Related Literature and Main Research Hypothesis

4.2.1 Related Literature

Initiated by the seminal papers of Becker (1968) and Allingham and Sandmo (1972), a sizeable and growing number of studies have addressed several aspects of tax evasion behavior.[3] However, our research question of whether individuals are more (or less) willing to evade taxes by underreporting income than by overdeducting negative income has not been answered so far. The only exceptions are Slemrod (2007)—discussed in the introduction—and Torgler (2013). The latter analyzes how moral suasion affects tax compliance behavior in a field experiment. In cooperation with the tax authority, he sends letters with normative appeals to taxpayers residing in a municipality in Switzerland and observes that moral suasion has only small effects on tax compliance. Apart from his main result, he finds evidence (although not statistically tested) for an asymmetric tax evasion behavior. In particular, he observes that people evade more taxes by deducting expenses and tax credits than by reporting income or wealth. This is in contrast to Slemrod (2007) who shows that underreporting income exceeds overreporting of deductions.

[3] As we are limited in presenting the whole body of literature on tax evasion and tax compliance, we recommend the papers of Alm et al. (1995), Andreoni et al. (1998), Torgler (2002), Hofmann et al. (2008), Alm (2012), and Slemrod (2019) for excellent literature reviews.

Because reporting positive and negative income have different effects on the current cash position,[4] the recent literature on reference point-dependent cheating behavior is closely connected to our study.[5] Although papers use a neutral non-tax context, these studies provide much evidence that tax evasion behavior may depend on whether subjects declare positive or negative income. Grolleau et al. (2016) conduct a laboratory experiment with a real-effort task. In their gain frame, subjects are paid according to their reported performance in this task. In their loss frame, subjects first receive the maximum possible amount, but they have to return money they do not earn with their reported performance. In both scenarios, the performance-based piece rate is the same. The results, however, indicate that subjects have a much higher willingness to cheat in the loss than in the gain frame, which can be explained by sufficient levels of loss aversion. Cameron et al. (2008) and Kern and Chugh (2009) find similar results and observe that behavior is more unethical when the outcomes are framed as losses than as gains.

If we transfer these results to a tax evasion context, we could hypothesize a higher tax evasion level in a situation where a taxpayer has to pay a tax due to positive income reported (perceptional loss frame) and a lower level in a situation where a taxpayer receives a tax refund due to negative income reported (perceptional gain frame). The large body of literature on the *income tax withholding phenomenon* supports these expectations.[6] This literature provides much evidence that individuals who pay too much tax in advance are more compliant than those who pay too few taxes beforehand. This withholding phenomenon can be explained by the value function of prospect theory (Kahneman and Tversky, 1979). If too much money as prepaid tax is withheld (overwithholding), individuals receive a tax refund at the year-end. This tax refund is regarded as a gain if the subject's reference point is its current cash position. Prospect theory predicts individuals will be risk averse in case of a gain frame. Thus, they become *more tax compliant* when confronted with a *tax refund*. In contrast, individuals who paid too few taxes in advance (underwithholding) have to pay additional taxes at the year-end. Hence, they view these *tax payments* as losses and become risk seeking, i.e., *less tax compliant*.

The first studies analyzing the withholding phenomenon experimentally are Robben et al. (1990a, 1990b). To create a realistic environment, a business situation was simulated in these experiments. Both studies reveal that tax evasion is higher if subjects have to pay an extra tax after insufficient taxes were withheld beforehand than if they receive a tax refund because too many taxes were paid in advance.

---

4 The two potential reference points most often discussed in the tax compliance literature are taxpayers' current and expected cash positions. Whereas in the former, tax payments (refunds) are framed as losses (gains), in the latter, tax payments (refunds) greater than expected are framed as losses (gains) and tax payments (refunds) less than expected are framed as gains (losses). As presented below, studies on the withholding phenomenon usually find evidence that taxpayers use their current cash position as a reference point. Because reporting additional positive income decreases the current cash position (tax payment) and reporting additional negative income increases it (tax refund), loss aversion can trigger an asymmetric tax evasion behavior if the tax payment is perceived as a loss and the tax refund as a gain.

5 Cheating and lying behavior is, for example, studied by Gneezy (2005), Sutter (2009), Erat and Gneezy (2012), Batigalli et al. (2013), Gneezy et al. (2013). Excellent literature reviews are presented by Ariely (2012) and Rosenbaum et al. (2014).

6 A literature overview is given by Hasseldine (1998a).

Furthermore, the authors find that a greater opportunity to evade taxes (e.g., a higher number of cash receipts or of business expense categories) increases tax evasion additionally. Jackson and Hatfield (2005) look at deduction choices and observe "that taxpayers who are due a tax refund prior to considering a judgemental tax deduction tend to select the conservative tax deduction, while taxpayers who owe additional taxes […] tend to select the aggressive tax deduction" (p. 148).[7] In other experimental studies, Martinez-Vazquez et al. (1992), Schepanski and Shearer (1995), and Copeland and Cuccia (2002) confirm the results on the withholding phenomenon. Chang and Schultz (1990), Elffers and Hessing (1997), and more recently Rees-Jones (2018) and Engström et al. (2015), provide empirical evidence for the phenomenon. For example, Rees-Jones (2018) shows that observed compliance behavior is consistent with the framing of tax payments as losses and tax refunds as gains. Engström et al. (2015) use data of Swedish taxpayers and observe that the probability of claiming deductions increases significantly for individuals who are in an underwithholding situation and have to pay taxes. They suggest that calibrating the tax law in such a way that most taxpayers receive tax refunds will increase tax compliance and will reduce auditing costs. Yaniv (1999) presents a theoretical analysis and shows that sufficiently high prepaid taxes may induce full compliance. Although there is already a large body of literature on tax evasion in the context of tax withholding, no focus has been placed on tax evasion of positive *and* negative income. Even though some studies distinguish between underreporting and overdeducting, they do not analyze the two tax evasion means separately. Robben et al. (1990a, 1990b) and Rees-Jones (2018), for example, measure tax evasion by underreporting and overdeducting, but they combine both factors into a single variable as their measure for tax evasion.

Please note that our study is related to the literature on the withholding phenomenon. However, our research question is different to that of this literature. The withholding literature focuses on the willingness to evade taxes when a taxpayer paid too much tax in advance (overwithholding) vs. when a taxpayer paid too few taxes beforehand (underwithholding). Through which channel the tax is evaded (positive or negative income) is not in focus. In our study, subjects are not confronted with an under- or overwithholding situation (i.e., individuals paid no taxes in advance). Instead we focus directly on the income reporting process and distinguish between both tax evasion strategies. This is why participants have to make two decisions: one tax evasion decision in case of positive income and one in case of negative income. Whereas the literature on the withholding phenomenon says something about the willingness to evade taxes in case of an under- or overwithholding situation, but does not analyze positive and negative income reporting separately, we focus on the tax evasion behavior when taxpayers are explicitly confronted with either positive or negative income.

Kirchler and Maciejovsky (2001) conduct a survey among self-employed and business entrepreneurs in Austria. In line with the literature on the withholding phenomenon, they show that while unexpected tax

[7] Jackson et al. (2005), Hatfield et al. (2008) and Jackson and White (2008) analyze the relation between taxpayers' prepayment positions and tax return preparation fees. These studies provide robust evidence that preparation fees are higher when taxpayers receive tax refunds than when they owe additional taxes.

payments reduce tax compliance, unexpected tax refunds increase tax compliance for self-employed taxpayers. In contrast, in case of business entrepreneurs, the authors observe that expected tax payments are associated with low and expected refunds with high tax compliance. Unexpected tax payments and unexpected tax refunds lead to a high level of tax compliance.[8] Although Kirchler and Maciejovsky are able to analyze whether subjects' tax compliance is higher when receiving a tax refund or when paying a tax, they cannot analyze whether tax evasion results from overdeduction or underreporting, as they use aggregated data. This is what our study links to.

4.2.2 Main Research Hypothesis

In our study, taxpayers have to make two decisions: one tax evasion decision in case of positive income (e.g., underreporting of business or nonbusiness income income) and one in case of negative income (e.g., overdeducting of expenses). The primary difference between both decisions is that subjects have to pay a tax in case of positive income and receive a tax refund in case of negative income. Although subjects are not confronted with an under- or overwithholding situation like in the previously discussed studies, they can—very similar to such situations—reduce a tax payment by declaring less positive income than they actually earned and they can increase a tax refund by declaring more negative income than they actually incurred. We therefore draw on the results of the literature on the income tax withholding phenomenon that finds that tax payments are associated with less tax compliance and tax refunds with more tax compliance.

Consequently, we hypothesize that subjects reveal a higher willingness to evade taxes in case of tax payments than in case of tax refunds. The rationale is as follows: Subjects decide on the change of the tax payment and on the change of the tax refund by their tax evasion decisions. If subjects evaluate the tax payment (that reduces their payoff) as a loss and the tax refund (that increases their payoff) as a gain (which is found by the withholding phenomenon literature), we expect them to be more risk seeking in the former than in the latter case. As a result, we hypothesize a higher tax evasion level for positive income (due to the tax payment) than for negative income (due to the tax refund). Our first and main research hypothesis is therefore:

Hypothesis 1: *Subjects evade more taxes in case of positive income than in case of negative income.*

A further explanation for an asymmetric tax evasion behavior is that individuals might not perceive both types of tax evasion as morally equivalent. Understating positive income implies paying less taxes to

[8] The general validity of these findings is limited, as Austrian legislation allows self-employed taxpayers to opt between two accounting principles (cash and accrual accounting), whereas it obliges entrepreneurs to use the more restrictive method (accrual accounting), which drives them to base their decisions on a long-term horizon. These different accounting methods can explain why the two groups use different cash positions and therefore show a diverse compliance behavior. Furthermore, the survey struggles with the problems of a self-selection bias and social desirability.

the tax authority and overstating negative income implies receiving more tax refunds from the tax authority. In the former individuals "keep their money" whereas in the latter individuals "take extra money" from the authority. If keeping own money is perceived as less immoral than taking money from others, we will expect more misreporting (and consequently more tax evasion) in the positive income than in the negative income scenario.

One reason for an asymmetric moral evaluation in this sense might be how acts of omission and acts of commission are morally evaluated. In the literature, there is robust evidence for the omission bias (Spranca et al., 1991; Ritov and Baron, 1995; Cox et al., 2017). In particular, individuals judge harmful omissions (e.g., failing to extinguish or report a fire) as less immoral than harmful commissions (e.g., igniting a fire). With respect to tax declaration decisions, failing to declare positive income to evade taxes and "keeping own money" is an act of omission whereas overreporting of expenditures and "taking extra money" is an act of commission. If taxpayers perceive acts of omissions as less immoral, tax evasion will be higher in case of positive income reporting than in case of negative income reporting.

A further rationale for an asymmetric moral evaluation builds on the general theory of "guilt aversion" developed by Battigalli and Dufwenberg (2007). The level of guilt perceived by an individual behaving immorally does not only depend on the immoral activity itself, but also on others' blame. If a taxpayer does not declare a positive income, she can claim that she simply forgot to declare. If a third person beliefs that this is true, the taxpayer is not really guilty from the third person's point of view and consequently guilt feelings are less pronounced for the taxpayer. However, if the taxpayer overstates expenditures (negative income), the third person will recognize this activity as a deliberately attempt to evade taxes and taxpayer's guilt feelings will be more pronounced. In line with this reasoning, "taking extra money" might also result in more others' blame and consequently more pronounced guilt feelings than "keeping own money". As tax evasion is consequently associated with less guilt feelings in the positive than in the negative income setting, tax evasion level should be higher in the former than in the latter.[9]

4.3 Experimental Design and Tax Evasion Measures

4.3.1 Experimental Design

We use three different experiments in our study to analyze tax evasion behavior in case of positive and negative income. In all experiments, the applied experimental design is as follows.[10] Subjects have to make tax evasion decisions for ten independent periods. In each of the ten periods, every participant receives an actual positive income PI_a *and* an actual negative income NI_a and decides how much of the positive *and* negative income she wants to report (PI_r and NI_r with $0 \leq PI_r \leq PI_a$ and $NI_r \leq NI_a \leq 0$).

[9] We thank the editor and an anonymous referee for pointing out that evading taxes by underreporting of positive income and overreporting of negative income might not be perceived as morally equivalent by taxpayers.

[10] The instructions of all experiments are displayed in Section 4.8 (Appendix A).

Actual positive and negative incomes are randomly drawn between 0 and 1,000 Lab-points and between 0 and -1,000 Lab-points, respectively, where only integer numbers appear.[11]

In case of positive income, a tax is raised with a rate τ of 50% based on the reported positive income PI_r, so that the tax payment T_{PI} can be calculated as $T_{PI} = PI_r \cdot \tau \geq 0$. With a probability of $p = 0.3$, the reported positive income is audited.[12] If the subject is caught cheating, she has to pay a penalty F_{PI} that is twice the evaded tax (i.e., $F_{PI} = 2 \cdot \tau \cdot (PI_a - PI_r) \geq 0$). If the subject is not caught cheating, no consequences occur. The subject's payoff (P_{PI}) in one period equals

$$P_{PI} = PI_a - PI_r \cdot \tau \geq 0 \tag{4.1}$$

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

$$P_{PI} = PI_a - PI_r \cdot \tau - 2 \cdot \tau \cdot (PI_a - PI_r) \geq 0 \tag{4.2}$$

In case of negative income, a tax refund is paid based on the reported negative income. The tax refund can be seen as a negative tax and is calculated as $T_{NI} = NI_r \cdot \tau \leq 0$. As we grant a complete and unrestricted loss offset, it holds that the higher the declared negative income, the higher the tax refund. The rules on audit probability ($p = 0.3$), tax rate ($\tau = 0.5$), and penalty ($F_{NI} = 2 \cdot \tau \cdot (NI_a - NI_r) \geq 0$) equal those in the case of a positive income. The payoff of one period for negative income (P_{NI}) is therefore given by

$$P_{NI} = NI_a - NI_r \cdot \tau \leq 0 \tag{4.3}$$

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

$$P_{NI} = NI_a - NI_r \cdot \tau - 2 \cdot \tau(NI_a - NI_r) \leq 0 \tag{4.4}$$

For both cases, Table 4.1 presents an example for the situation with or without an audit. Note that two payoffs are calculated separately in each period (one payoff for the positive and one for the negative income).[13] However, if an audit occurs, both the reported positive and the reported negative income are

[11] To abstract and simplify monetary values, we use Lab-points as currency units, where 1 Lab-point exactly corresponds to 0.01 Euro.

[12] Our chosen levels for the tax rate, audit probability, and fine are similar to those used in other tax evasion experiments (see Alm et al., 1995; Andreoni et al., 1998; Torgler, 2002; Hofmann et al., 2008; and Alm, 2012 for excellent literature reviews). Please notice that we tested the design of the Baseline ($PI_a = |NI_a|$) Experiment also with an audit probability of 0%. The results are robust to this variation (see section 4.6.1 for more details).

[13] In the instructions and in the entire experiment, subjects where confronted with positive values even in case of negative income. This has two main advantages. First, avoiding negative values reduces complexity for the participants. They do not have to enter a minus sign when declaring their negative income and they do not have to think in negative terms when making their decisions. Since subjects are usually not that familiar with negative values, we would possibly trigger an asymmetric tax evasion behavior otherwise. Second, applying positive values for economically negative income increases external validity since deductions, expenses, credits, and losses are usually entered as positive values in real tax returns (e.g., Form 1040). In order to ensure, that subjects understand that they have to enter a higher value than the actual value if they intend to evade taxes in case of negative income, we use an alert in the experiment. If a lower value is entered by a subject, the alert pops up and informs the subject that this is not allowed and a new input is required. Please note that we use positive values for positive income and negative values for negative income in the paper to better highlight that both kinds of income have two opposing economic consequences: positive income leads to a tax payment and negative income to a tax refund.

verified simultaneously. Therefore, either no penalty is raised at all (no audit occurs) or a penalty is raised for both the positive and negative income (audit occurs) if an individual evades taxes in both cases. If an individual evades only in one case and an audit occurs, the penalty is raised only for this case.

Table 4.1: Example

	positive income		**negative income**	
	no audit	**audit**	**no audit**	**audit**
actual income	600	600	-600	-600
reported income	350	350	-650	-650
tax	175	175	-325	-325
penalty	---	250	---	50
payoff	425	175	-275	-325

Note: For both income cases, this table presents an example for a situation with or without an audit.

At the beginning of the experiment, each subject is provided with an initial endowment of 10 Euro. The resulting (negative or positive) payoff from the experiment is offset against this initial endowment. To determine this payoff, the computer randomly decides which of the ten periods is relevant for the participant's payment at the end of the experiment. The payoff of the chosen period is converted into Euro and is then offset against the initial endowment of 10 Euro.[14] This total payment is paid out in cash immediately.

At the beginning of each experiment, the individuals are granted two training periods; these are not relevant for the final payment. After a subject decides how much of the positive and negative income she wants to report in one period (decision stage), the participant is informed about whether she was audited or not (information stage). Furthermore, she is provided a summary of information such as the actual and declared income, the tax payment/refund, the after-tax income, the penalty, and the payoff of the period for the positive and negative income. 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 their after-tax income for the situation with or without an audit. The experimental software was programmed and conducted with the software z-Tree (Fischbacher, 2007).

4.3.2 Tax Evasion Measures

To analyze the tax evasion behavior in our experiment, we use three different measures: (1) RATIO OF TAX EVASION FOR POSITIVE INCOME, (2) RATIO OF TAX EVASION FOR NEGATIVE

[14] It was not possible to obtain a negative payment, because of our parameter choice and the restriction of the reported negative income to $|NI_r| \leq |2 \cdot NI_a|$. This restriction ensures that the initial endowment is sufficient to compensate the highest potential negative income in the case with an audit for a negative income. Thus, even if the subject is audited, the payoff in a period never falls below -10 Euro. The total payment (after offsetting against the initial endowment) is, thus, at least zero Euro in all cases.

INCOME, and (3) RATIO OF TAX EVASION DIFFERENCE. We focus on relative measures because the amount for positive and negative income will usually differ. This enables us to compare the tax evasion behavior independently of the actual positive and negative income level. The RATIO OF TAX EVASION FOR POSITIVE (NEGATIVE) INCOME E^{PI} (E^{NI}) is calculated by dividing the difference of the ACTUAL and REPORTED POSITIVE INCOME (REPORTED and ACTUAL NEGATIVE INCOME) by the ACTUAL POSITIVE (NEGATIVE) INCOME:

$$E^{PI} = \frac{PI_a - PI_r}{PI_a} \geq 0 \quad \text{with} \quad PI_a \neq 0 \tag{4.5}$$

$$E^{NI} = \frac{NI_r - NI_a}{NI_a} \geq 0 \quad \text{with} \quad NI_a \neq 0 \tag{4.6}$$

Both measures equal zero if the income is declared truthfully, and both are greater than zero if an individual declares her income untruthfully. The *RATIO OF TAX EVASION FOR POSITIVE INCOME* is our measure for the willingness to evade taxes by *underreporting income*. The *RATIO OF TAX EVASION FOR NEGATIVE INCOME* is our measure for the willingness to evade taxes by *overdeducting expenses*.

The RATIO OF TAX EVASION DIFFERENCE accounts for the difference between both incomes in each period and is calculated by subtracting the RATIO OF TAX EVASION FOR POSITIVE INCOME from the RATIO OF TAX EVASION FOR NEGATIVE INCOME:

$$E^{difference} = E^{PI} - E^{NI} \tag{4.7}$$

4.4 Experiment 1 and 2: Baseline Experiments

4.4.1 Experiment Description

Treatment description. In our Baseline ($PI_a = |NI_a|$) experiment, a subject has to file two separate tax returns in each period: one for positive income and one for negative income (as described in Section 4.3.1). In this experiment, the absolute values of the actual positive and negative income are exactly equal to each other in one period (i.e., $PI_a = |NI_a|$).[15] This setting provides the highest level of control because a difference between tax evasion behavior in case of positive and negative income is not driven by different income levels in one period. Specifically, if tax evasion behavior depends on the pre-tax level, the ratios of tax evasion for positive and negative income cannot be compared so easily when the income level differ. However, if both amounts are identical in absolute terms, this effect cannot

[15] Note that the positive/negative income level is randomly drawn in each period and thus differs from period to period. However, in the Baseline ($PI_a = |NI_a|$) experiment, we ensure that the absolute amount of the actual positive and actual negative income are identical in one period.

bias the analysis. In our Section 4.9 (Appendix B), Figure 4.2 and Figure 4.3 give exemplary screenshots for the decision and information stage in this experiment.[16]

In the Baseline ($PI_a \neq |NI_a|$) experiment, the setup is completely identical to that of the first experiment. The only exemption is that the actual positive and negative incomes are now independently drawn from each other in every period. Consequently, the amount of positive income differs from the (absolute) amount of negative income (i.e., $PI_a \neq |NI_a|$). This design has two main advantages. First, because positive and negative income will usually differ in reality, this setting incorporates a higher external validity than the setting of the first experiment. Second, this experiment serves as a benchmark for the other two experiments in which identical positive and negative income in one period would trigger (by design) tax evasion behavior.

In all experiments, the computer randomly decides which of the ten periods is relevant for the participant's payment at the end of the experiment. In both Baseline experiments[17], each participant has to throw a six-sided die at the end of the experiment to decide whether the payoff of the positive or negative income is decisive. If the numbers 1, 2, or 3 occur, the payoff of the positive income determines the period payoff (otherwise the payoff of the negative income). Consequently, either the payoff of the positive or negative income is paid out (each with a probability of 50%), but not an aggregate of both payoffs.

Real world examples. The two baseline experiments are the most important experiments to isolate the asymmetric income reporting effect in the cleanest possible way. Nevertheless, they still have some practical implications. Our applied baseline setup imitates situations where positive and negative income reporting are not directly linked to each other. For example, a tax advisor reports income for two different clients or a taxpayer completes tax returns for different years. Moreover, the setup mirrors situations where an offset of positive and negative income is not allowed or is restricted. In almost all countries, for example, a complete and unrestricted general loss offset is not provided. Furthermore, some countries restrict the offsetting of positive and negative income for some income sources. For example in Germany, negative capital income (capital losses) cannot be offset against positive income from other income sources (e.g., business or nonbusiness income). Also the offset of domestic and foreign income is often restricted dependent on the national tax law and double taxation treaties between the involved countries. For example, if the exemption method is applied, negative foreign income cannot be offset against positive domestic income. Moreover, as positive and negative income reporting is

[16] To avoid that participants are overloaded or confused, we decided to fix the income's presentation on the screen in each period (i.e., that the positive income is presented on the left and the negative income on the right hand side). However, we reran the Baseline ($PI_a = |NI_a|$) experiment with a randomized presentation to exclude order effects. Our results are robust to this variation.

[17] As both experiments were run independently from each other (i.e., no random assignment of participants), we present them as two separate experiments and not as two different treatments.

unconnected in our baseline experiments, this setup might also mimic situations where only one type of income (either positive or negative) is reported.

4.4.2 Sample

Experiment 1 and 2 were conducted at the computerized experimental laboratory of the Leibniz University of Hanover with 59 student participants[18] (23 female and 36 male subjects) in total. Participants were recruited with hroot (Bock et al., 2014). Each individual participated in one experiment only (25 participants in Baseline ($PI_a = |NI_a|$) experiment and 34 participants in Baseline ($PI_a \neq |NI_a|$) experiment). Subjects earned on average 11.22 Euros in approximately 65 minutes (approximately 10.36 Euros per hour, no additional show-up fee was paid) with a minimum of 5.20 Euro and a maximum of 19.30 Euro. At the end of the experiment, participants are asked to answer a questionnaire which collects socio-demographic data and, for example, information on individual risk attitude and tax knowledge. In Section 4.10 (Appendix C), we provide an overview on the main characteristics of our participants in Table 4.8.

4.4.3 Results

For each of our three tax evasion measures, we calculated an average per participant over the ten periods. In the Baseline ($PI_a = |NI_a|$) experiment [Baseline ($PI_a \neq |NI_a|$) experiment], subjects evade 41.4% [39.3%] on average in case of positive income, whereas they only evade 27.7% [32.7%] in case of negative income. The difference between the positive and negative income ratios is 13.7% [6.6%] and is significant, with a p-value less than 0.05 in both experiments (Wilcoxon signed-rank test, two-tailed). This implies that, on average, subjects approximately evade 49.5% (= 41.4/27.7) [20.2% (= 39.3/32.7)] more taxes by underreporting income than by overdeducting expenses. In line with hypothesis 1, we can show that in both baseline experiments individuals evade more in case of positive income than in case of negative income.

To corroborate our descriptive and nonparametric results, we run linear regressions with all our observations (see Table 4.2). To take into account that subjects face repeated decision situations, we run OLS regressions with robust standard errors clustered at the subject level. In all models, the dependent variable is the RATIO OF TAX EVASION (for positive and negative income). The (independent) variable of interest is the dummy variable NEGATIVE INCOME which takes the value of 1 if the tax evasion decision was made for the negative income and 0 if the decision was made for the positive income (i.e., positive income scenario serves as the default). Thus, the coefficient of this dummy measures the difference between the ratio of tax evasion for positive and negative income.

[18] Please note that 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 experiments (Alm et al., 2015; Depositario et al., 2009; Liyanarachchi, 2007; Remus, 1996; Ashton and Kramer, 1980; Elliot et al., 2007).

Moreover, we consider a vector of experiment-specific and participant-specific variables. The following variables are included: LAST PERIOD AUDIT (1 if a subject was audited in the previous period, 0 otherwise); PERIOD (gives the current period in the experiment), ACTUAL POSITIVE INCOME < |ACTUAL NEGATIVE INCOME| ($PI_a < |NI_a|$) (1 if actual positive income < |actual negative income|, 0 otherwise); WHAT IF CALCULATIONS (number of "what if calculations", i.e., how often a subject used the computerized "what if" calculator in this period); FEMALE (female = 1, male = 0); ECONOMICS MAJOR (1 if the subject studies economics or management, 0 otherwise); BACHELOR (1 if the subject studies in a bachelor's degree program, 0 otherwise); AGE; RISK ATTITUDE (gives the subject's self-reported willingness to take a risk, measured on an 11-point scale where 0 = not willing to take risk and 10 = highly willing to take risk); INCOME (monthly income after fixed costs); and TAX KNOWLEDGE (gives subject's self-reported tax knowledge, measured on an 7-point scale where 1 = no knowledge and 7 = broad knowledge).

For both experiments, we observe a lower evasion level in case of negative income. In the Baseline ($PI_a = |NI_a|$) experiment (model 1) the coefficient is statistically significant at the 1%-level and in the Baseline ($PI_a \neq |NI_a|$) experiment (model 2) at the 5%-level. Therefore, regressions support our nonparametric results and hypothesis 1. The only further variable that is significant (at the 10%-level) in both models is FEMALE. The negative coefficient indicates that female participants evade less than male participants.[19]

[19] This is in line with other results that men tend to be more risk seeking than women (for a detailed survey of experimental and field data, see Eckel and Grossman, 2008; and Croson and Gneezy, 2009).

Table 4.2: OLS Regressions for our Two Baseline Experiments 1 and 2 (Dependent Variable: RATIO OF TAX EVASION)

	Baseline ($PI_a = \|NI_a\|$) Experiment	**Baseline ($PI_a \neq \|NI_a\|$) Experiment**
	Experiment 1	**Experiment 2**
	model 1	model 2
NEGATIVE INCOME	**-0.14*** **	**-0.07** **
	(0.05)	**(0.03)**
LAST PERIOD AUDIT	0.08	0.09**
	(0.05)	(0.04)
PERIOD	-0.00	0.01**
	(0.01)	(0.01)
ACTUAL POSITIVE INCOME < |ACTUAL NEGATIVE INCOME| ($PI_a < \|NI_a\|$)	---	-0.05
		(0.04)
WHAT IF CALCULATIONS	-0.01	0.03
	(0.01)	(0.02)
FEMALE	-0.18*	-0.14*
	(0.10)	(0.08)
ECONOMICS MAJOR	-0.09	0.05
	(0.13)	(0.07)
BACHELOR	-0.54***	0.09
	(0.11)	(0.09)
AGE	-0.04	0.01
	(0.02)	(0.01)
RISK ATTITUDE	0.04	0.02
	(0.02)	(0.03)
INCOME	0.00*	-0.00
	(0.00)	(0.00)
TAX KNOWLEDGE	-0.00	-0.00
	(0.04)	(0.04)
CONSTANT	1.58**	-0.03
	(0.65)	(0.37)
No. of Observations	500	680
No. of Clusters/Cubjects	25	34
R-squared	0.227	0.124

Note: In this table, the results of linear regressions are presented with RATIO OF TAX EVASION as dependent variable. To take into account that subjects face repeated decision situations, we run OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

4.5 Experiment 3

We conduct a third experiment with three treatments: one No-Offsetting Treatment and two Offsetting Treatments. The aim of this experiment is to test the robustness of our finding that individuals evade more by underreporting positive income than by overdeducting negative income with respect to different design modifications. We first create an environment in which an individual still makes two separate tax evasion decisions (as in both previous Baseline experiments) so that the decision stage remains identical for all presented experiments. However, in the Offsetting treatments the positive and negative incomes are now offset against each other. In both Baseline experiments either the payoff of the positive

or negative income was paid out. This creates a setting in which both decisions are clearly disconnected from each other. This disconnection is now removed in the two Offsetting treatments. Due to the resulting interaction of the positive and negative income reporting, it is arithmetically immaterial through which income channel (positive or negative) taxes are evaded in such a setting. Second, we vary the salience level of this offsetting (i.e., the intensity that the resulting interaction can be identified).

As illustrated in the example presented in Table 4.3, in the Offsetting-Low-Salience Treatment, both tax evasion decisions are clearly separated from each other, tax payment and tax refund as well as the potential penalties are separately determined for each income case, and the offsetting is realized at the end of a period by aggregating the resulting (after-tax and penalty) payoffs of the positive and negative income. Hence, in this treatment the offsetting of positive and negative income is not presented very saliently and the resulting interaction of income reporting is rather difficult to identify. In the Offset-High-Salience Treatment, positive and negative income are still declared separately, but an aggregation of both incomes follows immediately thereafter. The tax (tax payment or tax refund) and penalty are based on that aggregation. Thus, in this treatment the offsetting is presented very saliently and the interaction of income reporting can be identified very easily.

Table 4.3: Overview of the Offsetting Mechanisms in both Offsetting Treatments

Offsetting-Low-Salience Treatment			**Offsetting-High-Salience Treatment**	
	positive income	**negative income**		
actual income	600	-400	actual positive income	600
			actual negative income	-400
			actual pre-tax amount	200
declared income	350	-450	declared positive income	350
			declared negative income	-450
			declared pre-tax amount	-100
tax	175	-225	tax	-50
after-tax income	425	-175	after-tax income	250
penalty	250	50	penalty	300
yield of the period	175	-225		
yield of the period (total)	-50		yield of the period	-50

Note: This table presents an overview of the offsetting mechanisms in both Offsetting Treatments assuming an audit. Corresponding overviews are also presented to the participants in the information stages of the respective treatments (see Section 4.9, Appendix B, Figure 4.5 and Figure 4.7). Both Offsetting Treatments yield the same payoffs. The Offsetting-Low-Salience Treatment corresponds to the No-Offsetting Treatment with the exception that the yield of the period for the positive and negative incomes are accumulated at the end.

In all treatments, we apply the design of our Baseline ($PI_a \neq |NI_a|$) experiment for two main reasons. First, it seems to be more externally valid that positive and negative income differ. Second, if subjects are confronted with an identical actual positive and negative income in absolute terms as in the Baseline

($PI_a = |NI_a|$) experiment, the after-tax payoffs of positive and negative income will be identical (in absolute terms) in case of completely honest behavior. Because the payoffs of positive and negative income are aggregated at the end of a period in the Offsetting-Low-Salience and Offsetting-High-Salience experiments, the total payoff from one period would always be zero in both experiments in this case. Consequently, subjects would only earn the initial and constant endowment of 10 Euro. If subjects want to earn more than this endowment, they have to be dishonest. Thus, this design would immoderately trigger tax evasion. To avoid this bias, we decided using the setting of our Baseline ($PI_a \neq |NI_a|$) experiment in which the (absolute) amounts of positive and negative income differ.

4.5.1 No-Offsetting Treatment

The setup of the No-Offsetting Treatment is completely identical to the setup of our Baseline ($PI_a \neq |NI_a|$) experiment. However, we rerun this setup to be able to compare the results of this condition with the results of the two new Offsetting treatments. The No-Offsetting Treatment can also be seen as a replication of the Baseline ($PI_a \neq |NI_a|$) experiment. We therefore expect to find that individuals evade more taxes in case of positive income than in case of negative income (hypothesis 1).

4.5.2 Offsetting-Low-Salience Treatment

Treatment description. The Offsetting-Low-Salience Treatment differs from the No-Offsetting Treatment inasmuch as the after-tax payoffs of positive and negative income are now accumulated at the end of each period. Therefore, the individuals no longer have to throw the die at the end of the treatment in order to determine which income is relevant for their payoff. Instead, their payoff in one period results from the accumulation of the (after-tax) positive and negative income payoffs in this period. Thus, in contrast to the No-Offsetting Treatment, both decisions made in one period (declared positive and negative income) are decisive for the payout.[20] All other design parameters remain unchanged. Figure 4.4 and Figure 4.5 in Section 4.9 (Appendix B) give exemplary screenshots for the decision and information stages.

Although the after-tax payoffs are aggregated at the end of a period, each individual decides separately on how much she wants to declare as positive income and how much she wants to declare as negative income. Therefore, individuals make two separate decisions and are confronted both with paying a tax (in case of positive income) and receiving a tax refund (in case of negative income). Consequently, two separate taxable bases still exist in this setting (as in the Baseline experiments).

Formally, the aggregated payoff P is given by

$$P = P_{PI} + P_{NI} \tag{4.8}$$

[20] Even though the payoffs of after-tax positive and negative income are accumulated in one period, it is still valid that only one period is chosen randomly by the computer at the end of the experiment to determine the payment of each individual as in the previous experiments.

Compared to the No-Offsetting Treatment, we do not change the fiscal parameters (tax rate, penalty, and audit probability) or the audit process. If a participant is audited in a period, her declared positive and negative income are both verified as in the No-Offsetting Treatment or as in the previous Baseline experiments. Consequently, P_{PI} and P_{NI} presented in equations (4.1) to (4.4) are also valid for the Offsetting-Low-Salience Treatment. The subject's aggregated payoff in one period, therefore, equals

$$P = \underbrace{PI_a - PI_r \cdot \tau}_{P_{PI}} + \underbrace{NI_a - NI_r \cdot \tau}_{P_{NI}} \tag{4.9}$$

$$= PI_a + NI_a - \tau \cdot (PI_r + NI_r)$$

if no audit occurs and

$$P = \underbrace{PI_a - PI_r \cdot \tau - 2 \cdot \tau \cdot (PI_a - PI_r)}_{P_{PI}} + \underbrace{NI_a - NI_r \cdot \tau - 2 \cdot \tau \cdot (NI_a - NI_r)}_{P_{NI}} \tag{4.10}$$

$$= PI_a + NI_a - \tau \cdot (PI_r + NI_r) - 2 \cdot \tau \cdot (PI_a + NI_a - PI_r - NI_r)$$

if an audit occurs.

Hypothesis. In both Offsetting Treatments (with low and high salience), it is arithmetically immaterial through which income (positive or negative) taxes are evaded. In other words, it does not matter whether the individual evades taxes through underreporting positive income or by overdeducting negative income. Therefore, we should not observe a divergent tax evasion behavior. In the Offsetting-Low-Salience Treatment, however, the design is modelled in such a way that the presented positive income is clearly separated from the negative income. This implies that positive and negative incomes are kept isolated up to the aggregation and that the taxes and possible penalties are calculated individually for each income. Consequently, the offsetting of positive and negative income is not presented saliently and the resulting interaction of income reporting—that leads to the arithmetical irrelevance of whether taxes are evaded by underreporting positive or overdeducting negative income—is rather difficult to identify.

Salience effects are analyzed and proven in different tax and non-tax settings. For example, Rupert and Wright (1998) and Bordalo et al. (2012) study salience in risky decisions, Chetty et al. (2009), Finkelstein (2009), Bordalo et al. (2013), and Bordalo et al. (2016) in consumption decisions, Bordalo et al. (2015) in judicial decisions, Sausgruber and Tyran (2005, 2011) in voting decisions, Fochmann and Weimann (2013) in labor decisions. They all have in common that they provide clear evidence that decision maker's attention on an attribute of the decision task depends on the salience of this attribute. The higher the salience, the higher the attention and the processing of contained information.

In our setup, offsetting is one attribute determining the final payoff. However, offsetting is not presented saliently to the subjects in the Offsetting-Low-Salience Treatment. In line with the salience literature, we conjecture that this leads subjects to pay too little attention to the offsetting at the end of each period and we therefore expect them to decide on both tax evasion decisions independently. In other words, although the after-tax payoffs are aggregated at the very end, an individual might decide on the tax

evasion of positive income completely independently from her decision on negative income, as she is unaware of the subsequent aggregation when making her decision due to the low salience of offsetting. From this perspective, this creates the same decision environment as in the No-Offsetting Treatment. Thus, we should also observe a higher tax evasion level for positive than for negative income. Consequently, hypothesis 1 also applies here.

Real world examples. The setting in the Offsetting-Low-Salience Treatment mirrors tax reporting decisions where the positive and negative income reporting are clearly separated from each other (e.g., as different tax forms apply) and where the resulting interaction effects from reporting are not clearly linked to each other. In fact, taxpayers often report their income, deductions, and tax credits separately. In Germany, for example, taxpayers declare certain deductions (e.g., extraordinary expenses or expenses for household-related services) completely unrelated to their (positive) income. Furthermore, taxpayers have to use different tax forms for each income source (e.g., capital income, business income from self-employment, and income from employment).

As individuals make two separate tax evasion decisions for the positive and negative income in our Offsetting-Low-Salience Treatment, this setting can also be regarded as one taxpayer completing two separate tax returns. Although the taxpayer's wealth is affected by both tax returns in the end, the salience of the interaction of income reporting across the different tax returns is extremely low. Examples for such a scenario are a firm owner with two different corporations or a CEO managing two corporations with one company having generated a gain, whereas the other has generated a loss. Also parent and subsidiary companies, which are not organized as fiscal unity, have to complete separate tax returns. Although two taxpayers do exist from a legal perspective, the parent company can influence the subsidiary's management in such a way that they constitute only one taxpayer from an economic perspective. Furthermore, group reliefs and fiscal unities often underlie certain restrictions such as a minimum ownership rate or residential requirements. If these requirements are not met, the group relief or fiscal unity is not effective and negative income may not be offset against positive income within one consolidated tax return. Moreover, the U.K. offers a group relief which does not involve a combined tax return so that positive and negative income may be offset without demanding a common tax return.

4.5.3 Offsetting-High-Salience Treatment

Treatment description. In this treatment, each subject is again confronted with a positive and a negative income and has to declare her positive and negative incomes in each period (as before), but the declared positive and negative incomes are now aggregated to a single value that constitutes the base for the taxation (i.e., tax base equals $PI_r + NI_r$). Therefore, a subject is not confronted with paying a tax *and* receiving a tax refund, as in the previous treatments, but *either* has to pay a tax if the declared positive income is greater than the declared negative income *or* receives a tax refund if the declared negative income is greater than the declared positive income. Consequently, because only one common

tax base is calculated, there is also only one mutual possible penalty and one payoff of the period. All other design parameters remain unchanged. Figure 4.6 and Figure 4.7 in Section 4.9 (Appendix B) give exemplary screenshots for the decision and information stages.

Formally, the subject's payoff in one period equals

$$P = PI_a + NI_a - \tau \cdot (PI_r + NI_r) \tag{4.11}$$

if no audit occurs and

$$P = PI_a + NI_a - \tau \cdot (PI_r + NI_r) - 2 \cdot \tau \cdot (PI_a + NI_a - PI_r - NI_r) \tag{4.12}$$

if an audit occurs. Note that these payoff equations are identical to the corresponding equations 4.9 and 4.10 of the Offsetting-Low-Salience experiment. Again, it does not matter whether an individual evades taxes through underreporting positive income or overdeducting negative income. By creating this setting, we control whether the participant's tax evasion behavior is still divergent, although the arithmetical irrelevance of how to evade taxes is very salient now.

Hypothesis. In the Offsetting-High-Salience Treatment, the offsetting of positive and negative income is presented very saliently and the interaction of income reporting can be identified very easily (see Table 4.3). In particular, the tax and possible penalty are only calculated for one common taxable basis for positive and negative income, as they are accumulated immediately after the tax evasion decision. Thus, this setting clearly links both income reporting decisions and makes the interaction of positive and negative income declaration very salient. Compared to the Offsetting-Low-Salience Treatment, the salience of the offsetting is increased significantly. In line with the literature on salience effects, we conjecture that subjects now pay attention to the offsetting and the interaction of both income declarations leading them to decide on both incomes jointly. Consequently, we expect them to become aware of the arithmetical irrelevance of whether taxes are evaded by underreporting positive or overdeducting negative income and to decide on both incomes jointly by optimizing the declared pre-tax amount. We should therefore not observe a systematic difference between the tax evasion behavior for positive and negative income. Our second hypothesis therefore is:

Hypothesis 2: *In the Offsetting-High-Salience Treatment, the tax evasion behavior does not differ between the cases with positive and negative income.*

Real world examples. The setting in this treatment mirrors reporting decisions where positive and negative incomes are clearly linked to each other and where the resulting interaction effects can be identified easily. For example, if positive and negative income declaration is interpreted as business income and expense declaration, this setting reflects a situation where one taxpayer (an individual or a singly company) reports her income and expenses in one tax form (e.g., Form 1040 in the U.S.). Furthermore, if positive and negative incomes are materialized by different companies, this setting mirrors a tax regime that allow offsetting losses (i.e., negative income of one company) against gains

(i.e., positive income of another company). In the U.S., for example, commonly controlled corporations are allowed to file a consolidated tax return if certain requirements, such as 80% ownership of one company over the other, are fulfilled. Other countries, such as France, Germany, the Netherlands, and Spain also allow for fiscal unities which only have to file one consolidated tax return.

4.5.4 Sample

Experiment 3 was conducted at the computerized experimental laboratory of the University of Cologne with 146 participants (mostly undergraduates, 85 female and 61 male subjects) in total. Participants were recruited with ORSEE (Greiner, 2015). Participants were randomly assigned to one of our three treatments (39 participants in No-Offsetting Treatment, 39 in Offsetting-Low-Salience Treatment, 68 in Offsetting-High-Salience Treatment). Subjects earned on average 15.13 Euros in approximately 60 minutes (including a show-up fee of 4 Euros) with a minimum of 5.00 Euro and a maximum of 23.50 Euro. At the end of the experiment, participants are asked to answer a questionnaire which collects socio-demographic data and, for example, information on individual risk attitude and tax knowledge. In Section 4.10 (Appendix C), we provide an overview on the main characteristics of our participants in Table 4.8.

4.5.5 Results

As in our baseline experiments, we calculated an average per participant over the ten periods for each of our three tax evasion measures and for each treatment. Results can be found in Table 4.4 and Figure 4.1. In line with our previous findings and with hypothesis 1, we observe that individuals evade significantly more taxes in case of positive income than in case of negative income. This holds for the No-Offsetting and Offsetting-Low-Salience Treatment as well as when data is pooled over all three treatments. In the No-Offsetting Treatment, for example, we find that the RATIO OF TAX EVASION in case of positive income is 49.9%, but only 37.5% in case of negative income.

Although the level of the RATIO OF TAX EVASION DIFFERENCE decreases from 12.4% in the No-Offsetting Treatment to 5.7% in the Offsetting-Low-Salience Treatment, the difference in tax evasion in case of positive and negative income remains significant at the 5% level in this treatment (Wilcoxon signed-rank test, two-tailed). Hence, the modification that the payoffs of positive and negative income are aggregated so that both incomes are decisive for the subject's payment in each period does not cause the participants to change their divergent tax evasion behavior. On average, subjects evade approximately 11.0% (= 57.4/51.7) more taxes in case of positive income than in case of negative income. Overall, hypothesis 1 is supported again.

The only exemption is observed in the Offsetting-High-Salience Treatment. Here, the difference between positive and negative income is very small (3%) and not statistically significant anymore ($p = 0.1332$). Thus, we can conclude that in the Offsetting-High-Salience Treatment the observation that

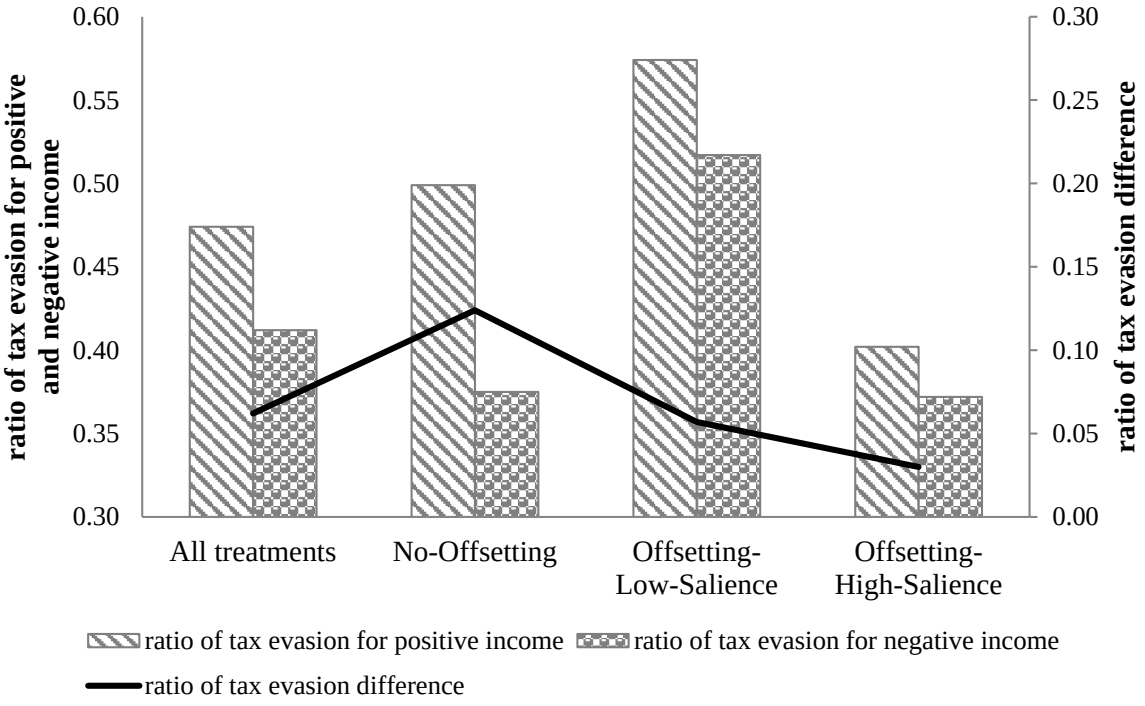
individuals significantly evade more taxes by underreporting income than by overdeducting expenses no longer persists. This supports hypothesis 2.

Table 4.4: Tax Evasion Behavior in our Experiment 3

Treatment	Statistics	Ratio of Tax Evasion for Positive Income (E^{PI})	Ratio of Tax Evasion for Negative Income (E^{NI})	Ratio of Tax Evasion Difference ($E^{difference}$)
all treatments (N = 146)	Mean	0.474	0.412	0.062
	Wilcoxon signed-rank test	$p < 0.001$		
No-Offsetting (N = 39)	Mean	0.499	0.375	0.124
	Wilcoxon signed-rank test	$p < 0.001$		
Offsetting-Low-Salience (N = 39)	Mean	0.574	0.517	0.057
	Wilcoxon signed-rank test	$p = 0.0190$		
Offsetting-High-Salience (N = 68)	Mean	0.402	0.372	0.030
	Wilcoxon signed-rank test	$p = 0.1332$		

Note: This table presents the descriptive statistics for the tax evasion behavior in all our treatments of experiment 3. The Wilcoxon signed-rank test (non-parametric test for dependent samples, two-tailed) analyzes whether the difference between the ratio of tax evasion for positive and negative income differs statistically in each treatment.

Figure 4.1: Tax Evasion Behavior in Experiment 3 (Means Presented)



To corroborate these descriptive and nonparametric results, we run OLS regressions with robust standard errors clustered at the subject level (see Table 4.5). We use the same procedure as we applied for our two baseline experiments (see Section 4.4.3 for all details). All regression results support our findings. In particular, we observe a significantly lower RATIO OF TAX EVASION FOR NEGATIVE INCOME than for POSITIVE INCOME over all treatments (model 1), in the No-Offsetting Treatment (model 2), and in the Offsetting-Low-Salience Treatment (model 3), but not in the Offsetting-High-Salience Treatment (model 4). The only further variable that is significant in all models is RISK ATTITUDE. The positive coefficient indicates that participants are evade more taxes the greater their level of risk seeking.

Table 4.5: OLS Regressions for our Experiment 3 (Dependent Variable: RATIO OF TAX EVASION)

	All Treatments	**No-Offsetting Treatment**	**Offsetting-Low-Salience Treatment**	**Offsetting-High-Salience Treatment**
	model 1	**model 2**	**model 3**	**model 4**
NEGATIVE INCOME	-0.06***	-0.13***	-0.06**	-0.03
	(0.01)	(0.03)	(0.02)	(0.02)
LAST PERIOD AUDIT	0.01	-0.00	0.05	-0.00
	(0.02)	(0.04)	(0.03)	(0.03)
PERIOD	0.01***	0.01***	0.01	0.01**
	(0.00)	(0.00)	(0.01)	(0.00)
ACTUAL POSITIVE INCOME < |ACTUAL NEGATIVE INCOME| ($PI_a < |NI_a|$)	-0.05**	-0.01	0.00	-0.10**
	(0.02)	(0.03)	(0.05)	(0.04)
WHAT-IF-CALCULATIONS	0.01*	0.03*	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
FEMALE	-0.06	-0.02	-0.06	-0.04
	(0.05)	(0.08)	(0.08)	(0.10)
ECONOMICS MAJOR	-0.05	-0.17*	-0.07	0.05
	(0.05)	(0.10)	(0.10)	(0.09)
BACHELOR	-0.09*	-0.07	-0.10	-0.17**
	(0.05)	(0.11)	(0.09)	(0.08)
AGE	-0.01*	-0.00	-0.01***	-0.00
	(0.00)	(0.01)	(0.00)	(0.01)
RISK ATTITUDE	0.06***	0.07***	0.07***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
INCOME	0.00	0.00**	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
TAX KNOWLEDGE	0.00	-0.01	-0.00	0.01
	(0.02)	(0.03)	(0.04)	(0.03)
CONSTANT	0.43***	0.28	0.51***	0.39*
	(0.13)	(0.26)	(0.18)	(0.22)
No. of Observations	2,819	759	780	1,280
No. of Clusters/Subjects	141	38	39	64
R-squared	0.173	0.254	0.230	0.136

Note: In this table, the results of linear regressions are presented with RATIO OF TAX EVASION as dependent variable. To take into account that subjects face repeated decision situations, we run OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

To compare our treatments, we also run OLS regressions with robust standard errors clustered at the subject level (see Table 4.6).[21] We compare our treatments with respect to all our three tax evasion measures. In addition, we are also interested in how much income is evaded in total (i.e., considering positive and negative income jointly). We focus on a relative measure again and calculate the RATIO OF TOTAL TAX EVASION as follows:

[21] As a robustness check, we rerun the regressions presented in Table 4.6 with the participant-averages (i.e., only one observation per participant is used in the regressions). All results are robust. All regression results can be found in Table 4.10 in Section 4.10 (Appendix C).

$$E^{Total} = \frac{PI_a - PI_r + NI_a - NI_r}{PI_a + |NI_a|} \geq 0 \qquad \text{with} \qquad PI_a + |NI_a| \neq 0 \tag{4.13}$$

whereas the numerator denotes the total evaded income and the denominator denotes the income that can be evaded maximally (in absolute terms).

In our OLS regressions, we use the following dependent variables: RATIO OF TAX EVASION DIFFERENCE (models 1 and 2), RATIO OF TAX EVASION FOR POSITIVE INCOME (models 3 and 4), RATIO OF TAX EVASION FOR NEGATIVE INCOME (models 5 and 6), and RATIO OF TOTAL TAX EVASION (models 7 and 8). As independent variables, we use treatment dummies. In models 1, 3, 5, and 7, we use the dummy "Offsetting Treatment" which takes the value of 1 if the decision was made in one of our two Offsetting Treatments (0 otherwise). In models 2, 4, 6, and 8, we use the dummies "Offsetting-Low-Salience Treatment" and "Offsetting-High-Salience Treatment" which take the value of 1 if the decision was made in the respective treatment. In all models, the No-Offsetting Treatment serves as the default. Thus, a coefficient of each dummy measures the difference between the respective condition and the No-Offsetting Treatment. In all models, experiment-specific and participant-specific variables are considered (not reported). All coefficients can be found in Table 4.9 in Section 4.10 (Appendix C).

The main results are as follows. First, RATIO OF TAX EVASION DIFFERENCE is significantly lower in the Offsetting Treatments than in the No-Offsetting Treatment (models 1 and 2). Consequently, offsetting significantly reduces the divergence between positive and negative income reporting. Second, this lower divergence is mainly driven by adjustments of positive income reporting. In fact, we do not observe significant treatment differences with respect to the RATIO OF TAX EVASION FOR NEGATIVE INCOME (models 5 and 6). However, with respect to the RATIO OF TAX EVASION FOR POSITIVE INCOME, we observe a lower level in the Offsetting Treatments (models 3 and 4). Third, TOTAL TAX EVASION does not differ significantly between treatments. Consequently, offsetting reduces the asymmetric income reporting, but does not reduce tax evasion in total.[22]

[22] We used Wald tests to check whether both coefficients of the Offsetting Treatments differ significantly in models 2, 4, 6 and 8. However, we did not observe any significant difference.

Table 4.6: Treatment Comparisons – OLS Regressions for our Experiment 3

Dependent Variable	**Ratio Of Tax Evasion Difference**		**Ratio Of Tax Evasion For Positive Income**		**Ratio Of Tax Evasion For Negative Income**		**Ratio Of Total Tax Evasion**	
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**	**model 7**	**model 8**
Offsetting Treatment	-0.10***		-0.12**		-0.02		-0.07	
	(0.03)		(0.06)		(0.05)		(0.05)	
Offsetting-Low-Salience Treatment		-0.09***		-0.07		0.02		-0.03
		(0.03)		(0.06)		(0.06)		(0.06)
Offsetting-High-Salience Treatment		-0.11***		-0.15**		-0.04		-0.10
		(0.03)		(0.06)		(0.06)		(0.06)
experiment-specific variables	yes	yes	yes	yes	yes	yes	yes	yes
participant-specific variables	yes	yes	yes	yes	yes	yes	yes	yes
CONSTANT	0.09	0.09	0.42***	0.44***	0.33**	0.35**	0.39***	0.41***
	(0.07)	(0.07)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
No. of Observations	1,409	1,409	1,410	1,410	1,409	1,409	1,410	1,410
No. of Clusters /Subjects	141	141	141	141	141	141	141	141
R-squared	0.043	0.043	0.193	0.198	0.164	0.167	0.203	0.207

Note: This table shows the results of OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

4.6 Robustness Checks

4.6.1 Fear of Further Losses

One further possible explanation for the lower tax evasion level in the case of negative income could be the participant's fear of further losses. For the negative income case, tax evasion that is detected due to an audit leads to even higher losses because of a penalty payment. Thus, individuals might be more tax compliant to prevent further losses. In order to analyze whether the fear of further losses is accountable for people evading less taxes by overdeducting negative income than by underreporting positive income, a No-Audit experiment based on our first experiment Baseline ($PI_a = |NI_a|$) is created and conducted with 17 participants. In contrast to our first experiment, the subject's tax declaration is not audited so that there are no possible penalty payments. If the fear of further losses causes the significantly different ratios of tax evasion in the cases of positive and negative income, this difference has to disappear in the No-Audit experiment.

As in our previous experiments, we calculated an average per participant over the ten periods for each of our three tax evasion measures. We find that TAX EVASION amounts to 63.3% in the positive

income case and to 51.7% in the negative income case. This difference is again highly significant with a p-value of 0.001 (Wilcoxon signed-rank test, two-tailed). Applying the same procedure as in the other experiments (see Section 4.4.3 for all details), we also run OLS regressions with robust standard errors clustered at the subject level. Results can be found in Table 4.11 (model 1) in Section 4.10 (Appendix C). Again, we find support for our hypothesis 1. In particular, we observe a significantly lower ratio of tax evasion for negative income than for positive income. Hence, the fear of further losses cannot be seen as the trigger for individuals to evade fewer taxes in the case of negative than in the case of positive income.

4.6.2 Rewording

In our experiments (and instructions), we decided to use the terminology "gain" for positive income and "loss" for negative income and "gain scenario" ("loss scenario") for the positive (negative) income scenario. The reason was to make it easier for our participants to clearly distinct between both scenarios and to be able to present positive values in both scenarios – so even in the negative income setting (see footnote 13 for a discussion).

To exclude that the use of this terminology has driven the observed asymmetric tax evasion behavior in case of positive and negative income, we conducted two robustness experiments. In particular, we rerun the Baseline ($PI_a \neq |NI_a|$) condition with a strong and the Offsetting-High-Salience condition without an asymmetric tax evasion effect. In these robustness experiments, we do not use the words "gain" and "loss" for positive and negative income. Instead, we use the terminology "revenue" (originally in German: Einnahme) and "expense" (originally in German: Ausgabe), respectively. 30 individuals participated in the Rewording-Baseline ($PI_a = |NI_a|$) experiment and 22 in the Rewording-Offsetting-High-Salience experiment.

For each of our three tax evasion measures, we calculated an average per participant over the ten periods again. In the Rewording-Baseline ($PI_a = |NI_a|$) experiment, participants evade 49.9% on average in case of positive income and 41.9% in case of negative income. The difference is significant with a p-value of 0.02 (Wilcoxon signed-rank test, two-tailed). In the Rewording-Offsetting-High-Salience experiment, we do not observe a significant difference between positive and negative income reporting (27.9% vs. 29.2%). We also run OLS regressions with robust standard errors clustered at the subject level. Results can be found in Table 4.11 (models 2 and 3) in Section 4.10 (Appendix C). Regression results support our descriptive findings. In particular, the coefficient of our NEGATIVE INCOME dummy is negative and significant at the 5%-level in the Baseline ($PI_a = |NI_a|$) condition, but insignificant in the Offsetting-High-Salience condition.

In line with our previous findings, we therefore find support for the asymmetric reporting in case of positive and negative income in the Baseline ($PI_a = |NI_a|$) condition again, but not in the Offsetting-High-Salience condition. Consequently, our results are robust to this variation.

4.6.3 Tax Evasion Differences over Time

We present the mean RATIO OF TAX EVASION DIFFERENCE separately for each period in Table 4.7 to analyze whether subjects experience some kind of learning in the course of the experiment. Regarding both Baseline experiments and the No-Offsetting Treatment, we find that subjects nearly constantly underreport more positive income than they overdeduct negative income. While the mean RATIO OF TAX EVASION DIFFERENCE in the Baseline ($PI_a = |NI_a|$) experiment and No-Offsetting Treatment is similar for all periods and for the last five periods, the ratio is even higher in the last five periods in the Baseline ($PI_a \neq |NI_a|$) experiment. Consequently, we do not find any evidence for learning effects so that we can conclude a systematic and persisting asymmetric tax evasion behavior.

Analyzing the Offsetting-Low-Salience Treatment, we also find that the mean RATIO OF TAX EVASION DIFFERENCE is smaller for all periods (5.72%) than for the last five periods (7.47%). Hence, even for the Offsetting-Low-Salience Treatment where one might have expected to observe learning effects (because of the irrelevance of how to evade taxes), we do not find the systematic positive RATIO OF TAX EVASION DIFFERENCE to disappear. In the Offsetting-High-Salience Treatment, we observe a nearly constant RATIO OF TAX EVASION DIFFERENCE at a relatively low level. Consequently, it does not seem that learning effects drive our results.

Table 4.7: Mean Ratio of Tax Evasion Difference by Period

	Experiment 1	**Experiment 2**	**Experiment 3**		
Period	**Baseline ($PI_a = \|NI_a\|$) experiment**	**Baseline ($PI_a \neq \|NI_a\|$) experiment**	**No-Offsetting Treatment**	**Offsetting-Low-Salience Treatment**	**Offsetting-High-Salience Treatment**
1	12.56%	6.79%	11.74%	1.64%	4.10%
2	13.59%	8.97%	12.91%	-4.53%	5.79%
3	16.81%	1.47%	11.34%	6.88%	4.33%
4	13.46%	-2.29%	10.91%	8.33%	-1.34%
5	18.60%	-1.84%	20.56%	7.53%	3.57%
6	17.86%	15.53%	8.50%	9.48%	0.75%
7	4.88%	14.63%	12.56%	6.90%	3.15%
8	1.20%	7.84%	5.58%	1.23%	3.17%
9	22.30%	6.70%	15.35%	14.54%	2.41%
10	16.44%	8.68%	13.87%	5.18%	3.58%
mean (all Periods)	13.77%	6.65%	12.33%	5.72%	2.95%
mean (last five Periods)	12.54%	10.68%	11.17%	7.47%	2.61%

Note: In this table, we present the mean RATIO OF TAX EVASION DIFFERENCE by period. The mean RATIO OF TAX EVASION DIFFERENCE for all periods and for the last five periods are presented in the last two columns.

4.6.4 Bunching

In our experiments, the limit for actual negative income is set to 1,000. Although the audit probability is 30% independent of the reported income, it might be that participants' subjective probability of an audit increases if they report negative income above this limit. To exclude that bunching at or below 1,000 has driven the observed asymmetric tax evasion behavior in case of positive and negative income, we rerun our OLS regressions for all three experiments.[23] The new regression results can be found in Table 4.12 in Section 4.10 (Appendix C).

First, we excluded all periods in which reported negative income equals 1,000 (models 1 to 3). This excludes all observations where a bunching at 1,000 has possibly taken place. Second, we excluded all periods in which reported negative income is between 900 and 1,000 (models 4 to 6). This excludes all observations where a bunching at and below 1,000 has possibly taken place. In all regression analyses our previous results are robust. In particular, we observe that TAX EVASION is significantly lower for negative income than for positive income. Consequently, bunching has not driven the observed asymmetric income reporting.

4.7 Discussion

The main findings of our study are threefold. First, we find that subjects are systematically more willing to evade taxes by underreporting positive income (e.g., business or nonbusiness income) than by overdeducting negative income (e.g., deductions, credits, or losses). Second, offsetting of positive and negative income reduces the asymmetric tax evasion behavior. Third, the salience of income interaction plays an important role in tax evasion decisions. This emphasizes that the design of tax returns can have a serious influence on tax evasion behavior.

As revenue bodies are increasingly confronted with reduced audit capacities[24], the decisions on whom to audit and what to audit become essential to ensure that tax evasion is combatted effectively. If both tax evasion means are assumed to be substitutes, our results indicate that tax authorities should pay more attention to the underreporting of income than to the overreporting of credits and offsets to income (which is in line with Slemrod, 2007). Our results further suggest that the resulting divergent tax evasion behavior seems to be prominent when the salience of income interaction is low (e.g., for taxpayers with income from different sources who have to complete an own tax form for each income source). As both tax evasion means are usually no perfect substitutes in reality, our identified divergence effect adds to

[23] In particular, we rerun model 1 from Table 4.2 for experiment 1, model 2 from Table 4.2 for experiment 2, and model 1 from Table 4.5 for experiment 3.

[24] The IRS is currently afflicted with monetary cuts, resulting in fewer audits and lower revenues collected. In 2015, the Commissioner of Internal Revenue, John A. Koskinen, stated that Fiscal Year (FY) 2014 "marked the fourth consecutive year IRS appropriations were reduced, which required us to change our approach to taxpayer service and enforcement operations. […] During FY 2014, the IRS audited tax returns of about 1.2 million individuals, which is nearly 12 percent less than the previous year and the lowest number since FY 2005. We estimate that, as a result of these enforcement cuts, the government will lose at least $2 billion in revenue that otherwise would have been collected" (Internal Revenue Service, 2015, p. iii).

the monetary effects induced by the different benefits and costs of both evasion opportunities. Future research is needed to examine the strength of each effect.

With our study we generate four contributions to the academic literature. First, we contribute to the general tax literature by showing that the willingness to evade taxes depends on the sort of income reported. Obviously, taxes can be evaded by both underreporting of positive income (e.g., business income) and by overdeduction of negative income (e.g., expenses, credits, losses). However, there is only rare empirical evidence on whether tax evasion behavior differs significantly between underreporting and overdeduction. By showing that taxpayers report positive and negative income asymmetrically, our study provides new evidence to this literature. As we find that subjects evade more in case of tax payments than in case of tax refunds, our study is also related to the literature on the income tax withholding phenomenon.

Second, by using controlled laboratory experiments we are able to cleanly identify causal effects which is a challenging issue when using observational or archival tax data. In this regard, we also contribute to the experimental and behavioral literature on tax compliance (e.g., Torgler, 2002; Kirchler et al., 2008; Alm and Torgler, 2011; Alm, 2012; Kogler et al., 2013; Blaufus et al., 2016). This strand of literature studies the foundations of tax evasion to understand how taxpayers can be influenced by tax regulations to enhance tax acceptance and to reduce tax evasion. If a better understanding helps to combat tax evasion more effectively, social welfare might increase (e.g., when tax authority's enforcement costs are reduced or when additional tax revenue is used to increase public good provision).

Third, our paper contributes to the general experimental and behavioral literature studying the influence of framing effects on the cheating and lying behavior of individuals (e.g., Grolleau et al., 2016). In line with prospect theory (Kahneman and Tversky, 1979), this literature provides clear evidence that subjects have a much higher willingness to cheat in a loss than in a gain frame. We contribute to this strand of literature by showing that individuals reveal a higher willingness to evade taxes in a loss frame (i.e., when confronted with a tax payment that reduces their payoff) than in a gain frame (i.e., when confronted with receiving a tax refund that increases their payoff).

Fourth, our study contributes to the literature on salience effects (e.g., Chetty et al., 2009; Bordalo et al., 2012; Sausgruber and Tyran, 2005, 2011). Although salience effects are shown to be important in different tax and non-tax settings, we are (to our knowledge) the first who study salience effects in a tax evasion context explicitly. In particular, we show that the observed asymmetric tax evasion behavior substantially depends on the salience of the income reporting interaction.

4.8 Appendix A: Experimental Instructions

We divided the instructions into different parts. The beginning part is identical in all experiments, whereas the following parts differ between the experiments. In the following, the instructions (originally written in German) are presented.

A1 Beginning Instructions of All Experiments

Thank you for participating at the today's experiment. For your participation you receive 10 Euros in advance (starting capital). Your overall earnings can either increase or decrease in the course of the experiment. How much you earn in total depends on your decisions and on chance. This instruction elucidates how you may influence the money you earn in this experiment by your decisions. Thus, read carefully throughout the following paragraphs.

We would like to inform you that you are not allowed to communicate with other participants or leave your seat throughout the whole experiment. If you have questions please raise your hand. We will come up to you to answer your questions.

For reasons of simplification we will not calculate with Euro-amounts in the experiment, but with lab-points. Thereby 1 lab-point exactly corresponds to 1 Euro-cents. That means 100 lab-points exactly correspond to 1 Euro.

The experiment consists of 10 periods in total which are independent from each other. At the end of the experiment one period is randomly drawn that determines your payoff.

A2 Specific Instructions of the Baseline Experiments

[*Baseline ($PI_a = |NI_a|$) experiment only:*

1. Gain and Loss Scenario

In every period you are assigned an integer amount X. This amount is randomly drawn by the computer and may take on integer numbers between 0 and 1000 lab-points. From one period to another the amount can differ and will be displayed to you before every single decision.

There is either a gain or a loss scenario in every period. However only one of those amounts is relevant for the payoff the end of the experiment. Which of both scenarios is present is not known to you before your decision but dependent on chance. The gain and the loss scenario both occur with a probability of 50%, respectively. In the gain scenario you are assigned a pre-tax-gain of X lab-points. In the loss scenario you are assigned a pre-tax-loss of X lab-points.

Please remember: In both scenarios the amount X is positive (thus greater than zero), but in the gain scenario the amount concerns a gain and in the loss scenario it concerns a loss. Therefore it holds that

your starting capital of 10 Euros is increased in the gain scenario but decreased in the loss scenario. How your starting capital exactly changes is described hereafter.]

[*Baseline ($PI_a \neq |NI_a|$) experiment only:*

1. Gain and Loss Scenario

In every period you are assigned a pre-tax-gain as well as a pre-tax-loss. The respective amounts of the pre-tax-gain and pre-tax-loss are randomly drawn by the computer and may take on integer numbers between 0 and 1000 lab-points. As the computer independently draws the amounts of the pre-tax-gain and pre-tax-loss, these two amounts can differ in height. Furthermore, it pertains: From one period to another the amount of the pre-tax-gain as well as of the pre-tax-loss can differ and will be displayed to you before every single decision.

You are assigned one pre-tax-gain and one pre-tax-loss in every period. However only one of those amounts is relevant for the payoff the end of the experiment. Whether the gain or the loss scenario is relevant for your payoff is not known to you before your decision but is dependent on chance. The gain and the loss scenario both occur with a probability of 50%, respectively.

Please remember: In both scenarios the amount is positive (thus greater than zero), but in the gain scenario the amount concerns a gain and in the loss scenario it concerns a loss. Therefore it holds that your starting capital of 10 Euros is increased by a gain but decreased by a loss. How your starting capital exactly changes is described hereafter.]

2. Tax Declaration

In every period there is a tax with a tax rate of 50%. The amount of the tax is assessed according to your pre-tax-amount that you are asked to declare for the gain scenario as well as for the loss scenario. Hereto, you just assess how much of the actual pre-tax-gain and of the pre-tax-loss you want to declare. Only integer values are possible to be declared. Please remember that the declaration of your pre-tax-gain is independent of the declaration of your pre-tax-loss. Thus, both declarations can deviate from each other. (Please note: The declaration of your pre-tax-gain as well as of your pre-tax-loss are the only two decisions that you have to take in a period.)

The following constraint holds: Your declared pre-tax-gain may not exceed your actual pre-tax-gain, but may also not be smaller than zero. Your declared pre-tax-loss may not deceed your actual pre-tax-loss, but may also not exceed twice as much as your actual pre-tax-loss.

Please remember: The tax is due in both, the gains and the loss scenario. However, the effect of the tax is a bit different: In the gain scenario you have to pay a tax so that your gain is decreasing. In the loss scenario you receive a tax refund so that your loss decreases.

The tax payment and the tax refund are thus calculated the following:

In the gain scenario: tax payment = 0.5 x *declared* pre-tax-gain

In the loss scenario: tax refund = 0.5 x *declared* pre-tax-loss

Your after-tax-gain and after-tax-loss are thus calculated the following:

In the gain scenario: after-tax-gain = *actual* pre-tax-gain - tax payment

In the loss scenario: after-tax-loss = *actual* pre-tax-loss - tax refund

3. Audit of your tax declaration

With a probability of 30% both of your declarations on your pre-tax-gain and pre-tax-loss are audited. With a probability of 70% your declarations are not audited. If you are audited and the actual and declared pre-tax-gain or pre-tax-loss do not coincide, a penalty is charged. The penalty amounts to twice the evaded tax in the gain scenario and twice the overpaid obtained tax refund in the loss scenario:

In the gain scenario: penalty = 2 x evaded tax

In the loss scenario: penalty = 2 x overpaid obtained tax refund

Thereby it holds:

evaded tax = 0.5 x (*actual* pre-tax-gain - *declared* pre-tax-gain)

overpaid obtained tax refund = 0.5 x (*declared* pre-tax-gain - *actual* pre-tax-gain)

Please remember: As either the gain or the loss scenario is present the penalty only has to be paid once for the respective applicable scenario. Please also remember: If the declared and actual pre-tax-amount coincide no penalty is charged as the difference of actual and declared pre-tax-amount is zero.

Your yield of the period in case of an *audit* is thus calculated the following:

In the gain scenario: yield of the period (gain) = after-tax-gain - penalty

In the loss scenario: yield of the period (loss) = after-tax-loss + penalty

Please remember: In the gain scenario the penalty results in a decrease of the gain. In the loss scenario the penalty results in an increase of the loss.

Your yield of the period in case of *no audit* is thus calculated the following:

In the gain scenario: yield of the period (gain) = after-tax-gain

In the loss scenario: yield of the period (loss) = after-tax-loss

After every period you are informed whether you have been audited or not. Furthermore, you are granted an overview of all important values as well as of your yield of the period in the gain and loss scenario.

4. What if calculator

For both decisions on how much of the pre-tax-gain and pre-tax-loss you want to declare, you have the possibility to perform what if calculations on the computer (bottom screen). For this purpose enter the pre-tax-gain and pre-tax-loss that you wish to declare. Afterwards the yield of the period for the gain and the loss scenario is announced to you for the case with and without an audit. Please remember, that what if calculations are not relevant for your payoff of the experiment.

In addition you can use the pocket calculator that is provided at your site for own calculations.

5. Total payoff from the experiment

After you have made your decisions in all 10 periods *one* period is randomly drawn by the computer at the end of the experiment and presented to you on the screen. To determine whether there is a gain or loss scenario in this period you are asked to throw a six-sided dice at the experimenters' desk. If you dice a 1, 2 or 3 you are in a gain scenario, if you dice a 4, 5 or 6 you are in a loss scenario. The yield of the period that resulted in the respective period for the diced scenario is converted into Euro and reckoned up with your starting capital of 10 Euro. If there is a gain scenario your starting capital increases by the amount. If there is a loss scenario your starting capital decreases by the amount. The resulting total payoff is cashed out to you subsequent to the experiment.

Please remember: It is ensured that you may never sustain any loss after your starting capital is reckoned up with the yield of the period.

6. Training periods

Bevor the real experiment with 10 periods starts there is a rehearsal with 2 training periods. The decisions you make in these training periods have no influence on the payoff of the experiment.

A3 Specific Instructions of the Offsetting-Low-Salience Treatment

1. Pre-Tax-Gain and Pre-Tax-Loss

In every period you are assigned a pre-tax-gain as well as a pre-tax-loss. The respective amounts of the pre-tax-gain and pre-tax-loss are randomly drawn by the computer and may take on integer numbers between 0 and 1000 lab-points. As the computer independently draws the amounts of the pre-tax-gain and pre-tax-loss, these two amounts can differ in height. Furthermore, it pertains: From one period to another the amount of the pre-tax-gain as well as of the pre-tax-loss can differ and will be displayed to you before every single decision.

Please remember: In both scenarios the amount is positive (thus greater than zero), but in the gain scenario the amount concerns a gain and in the loss scenario it concerns a loss. Therefore it holds that your starting capital of 10 Euros is increased by a gain but decreased by a loss. How your starting capital exactly changes is described hereafter.

2. Tax Declaration

In every period there is a tax with a tax rate of 50%. The amount of the tax is assessed according to your pre-tax-amount that you are asked to declare for the gain scenario as well as for the loss scenario. Hereto, you just assess how much of your actual pre-tax-gain and of your pre-tax-loss you want to declare. Only integer values are possible to be declared. Please remember that the declaration of the pre-tax-gain is independent of the declaration of the pre-tax-loss. Thus, both declarations can deviate from each other. (Please note: The declaration of your pre-tax-gain as well as of your pre-tax-loss are the only two decisions that you have to take in a period.)

The following constraint holds: Your declared pre-tax-gain may not exceed your actual pre-tax-gain, but may also not be smaller than zero. Your declared pre-tax-loss may not deceed your actual pre-tax-loss, but may also not exceed twice as much as your actual pre-tax-loss.

Please remember: The tax is due in both, the gains and the loss scenario. However, the effect of the tax is a bit different: In the gain scenario you have to pay a tax so that your gain is decreasing. In the loss scenario you receive a tax refund so that your loss decreases.

The tax payment and the tax refund are thus calculated the following:

In the gain scenario: tax payment = 0.5 x *declared* pre-tax-gain

In the loss scenario: tax refund = 0.5 x *declared* pre-tax-loss

Your after-tax-gain and after-tax-loss are thus calculated the following:

In the gain scenario: after-tax-gain = *actual* pre-tax-gain - tax payment

In the loss scenario: after-tax-loss = *actual* pre-tax-loss - tax refund

3. Audit of your tax declaration

With a probability of 30% both of your declarations on your pre-tax-gain and pre-tax-loss are audited. With a probability of 70% your declarations are not audited. If you are audited and the declared and actual pre-tax-gain or pre-tax-loss do not coincide, a penalty is charged. The penalty amounts to twice the evaded tax in the gain scenario and twice the overpaid obtained tax refund in the loss scenario:

In the gain scenario: penalty = 2 x evaded tax

In the loss scenario: penalty = 2 x overpaid obtained tax refund

Thereby it holds:

evaded tax = 0.5 x (*actual* pre-tax-gain - *declared* pre-tax-gain)

overpaid obtained tax refund = 0.5 x (*declared* pre-tax-gain - *actual* pre-tax-gain)

Please remember: If the declared and actual pre-tax-amount coincide no penalty is charged as the difference of actual and declared pre-tax-amount is zero.

Your yield of the period in case of an *audit* is thus calculated the following:

In the gain scenario: yield of the period (gain) = after-tax-gain - penalty

In the loss scenario: yield of the period (loss) = after-tax-loss + penalty

Please remember: In the gain scenario the penalty results in a decrease of the gain. In the loss scenario the penalty results in an increase of the loss.

Your yield of the period in case of *no audit* is thus calculated the following:

In the gain scenario: yield of the period (gain) = after-tax-gain

In the loss scenario: yield of the period (loss) = after-tax-loss

After every period you are informed whether you have been audited or not. Furthermore, you are granted an overview of all important values as well as of your yield of the period in the gain and loss scenario.

4. Yield of the period (total):

Your aggregated yield of the period, the yield of the period (total), is calculated the following:

Yield of the period (total) = yield of the period (gain) - yield of the period (loss)

This amount is relevant for the payoff at the end of the experiment.

5. What if calculator

For both decisions on how much of the pre-tax-gain and pre-tax-loss you want to declare, you have the possibility to perform what if calculations on the computer (bottom screen). For this purpose enter the pre-tax-gain and pre-tax-loss that you wish to declare. Afterwards the yield of the period is announced to you for the case without and with an audit. Please remember, that what if calculations are not relevant for your payoff of the experiment.

In addition you can use the pocket calculator that is provided at your site for own calculations.

6. Total payoff from the experiment

After you have made your decisions in all 10 periods *one* period is randomly drawn by the computer at the end of the experiment and presented to you on the screen. The yield of the period (total) that resulted in the respective period is converted into Euro and reckoned up with your starting capital of 10 Euro. The resulting total payoff is cashed out to you subsequent to the experiment.

Please remember: It is ensured that you may never sustain any loss after your starting capital is reckoned up with the yield of the period (total).

7. Training periods

Bevor the real experiment with 10 periods starts there is a rehearsal with 2 training periods. The decisions you make in these training periods have no influence on the payoff of the experiment.

A4 Specific Instructions of the Offsetting-High-Salience Treatment

1. Gain and Loss

In every period you are assigned a gain as well as a loss. The respective amounts of the gain and loss are randomly drawn by the computer and may take on integer numbers between 0 and 1000 lab-points. As the computer independently draws the amounts of the gain and loss, these two amounts can differ in height. Furthermore, it pertains: From one period to another the amount of the gain as well as of the loss can differ and will be displayed to you before every single decision.

The difference between actual gain and actual loss results in the actual pre-tax-amount:

actual pre-tax-amount = *actual* gain - *actual* loss

Please remember: As the actual gain can be both, greater as well as smaller, than the actual loss the actual pre-tax-amount can be positive as well as negative.

2. Tax Declaration

In every period there is a tax with a tax rate of 50%. The amount of the tax is assessed according to your declared pre-tax-amount that you are asked to declare. Hereto, you just assess how much of your actual gain and loss you want to declare. Only integer values are possible to be declared. Please remember that the declaration of the gain is independent of the declaration of the loss. (Please note: The declaration of your gain as well as of your loss are the only two decisions that you have to take in a period.)

The declared pre-tax-amount is calculated the following:

declared pre-tax-amount = *declared* gain - *declared* loss

Please remember: As the declared gain can be both, greater as well as smaller, than the declared loss the declared pre-tax-amount can be positive as well as negative.

The following constraint holds: Your declared gain may not exceed your actual gain, but may also not be smaller than zero. Your declared loss may not deceed your actual loss, but may also not exceed twice as much as your actual loss.

The tax amounts to 50% of your declared pre-tax-amount, that means:

tax = 0.5 x *declared* pre-tax-amount

Your after-tax-amount is thus calculated the following:

after-tax-amount = *actual* pre-tax-amount - tax

Please remember: The tax is due for both, a positive as well as a negative declared pre-tax-amount. However the effect of the tax is a bit different: For a positive pre-tax amount the tax is positive. That means you have to pay a tax and your after-tax-amount decreases. For a negative pre-tax amount the tax is negative. That means you receive a tax refund and your after-tax-amount increases.

3. Audit of your tax declaration

With a probability of 30% the declaration on your declared pre-tax-amount is audited. With a probability of 70% your declaration is not audited. If you are audited and the declared and actual pre-tax-amount does not coincide, a penalty is charged. The penalty amounts to twice the evaded tax:

penalty = 2 x evaded tax

Thereby it holds:

evaded tax = 0.5 x (*actual* pre-tax-amount - *declared* pre-tax-amount)

Please remember: If the declared and actual pre-tax-amount coincide no penalty is charged as the difference of actual and declared pre-tax-amount is zero.

Your yield of the period in case of an *audit* is thus calculated the following:

yield of the period = after-tax-amount - penalty

Your yield of the period in case of *no audit* is thus calculated the following:

yield of the period = after-tax-amount

After every period you are informed whether you have been audited or not. Furthermore, you are granted an overview of all important values as well as of your yield of the period.

Please remember that your yield of the period can also be negative. In that case, your starting capital decrease by this amount. If the yield of the period is positive your starting capital increases by this amount.

4. What if calculator

For both decisions on how much of the gain and loss you want to declare, you have the possibility to perform what if calculations on the computer (bottom screen). For this purpose enter the gain and loss that you wish to declare. Afterwards the yield of the period is announced to you for the case without and with an audit. Please remember, that what if calculations are not relevant for your payoff of the experiment.

In addition you can use the pocket calculator that is provided at your site for own calculations.

5. Total payoff from the experiment

After you have made your decisions in all 10 periods *one* period is randomly drawn by the computer at the end of the experiment and presented to you on the screen. The yield of the period that resulted in the respective period is converted into Euro and reckoned up with your starting capital of 10 Euro. The resulting total payoff is cashed out to you subsequent to the experiment.

Please remember: It is ensured that you may never sustain any loss after your starting capital is reckoned up with the yield of the period.

6. Training periods

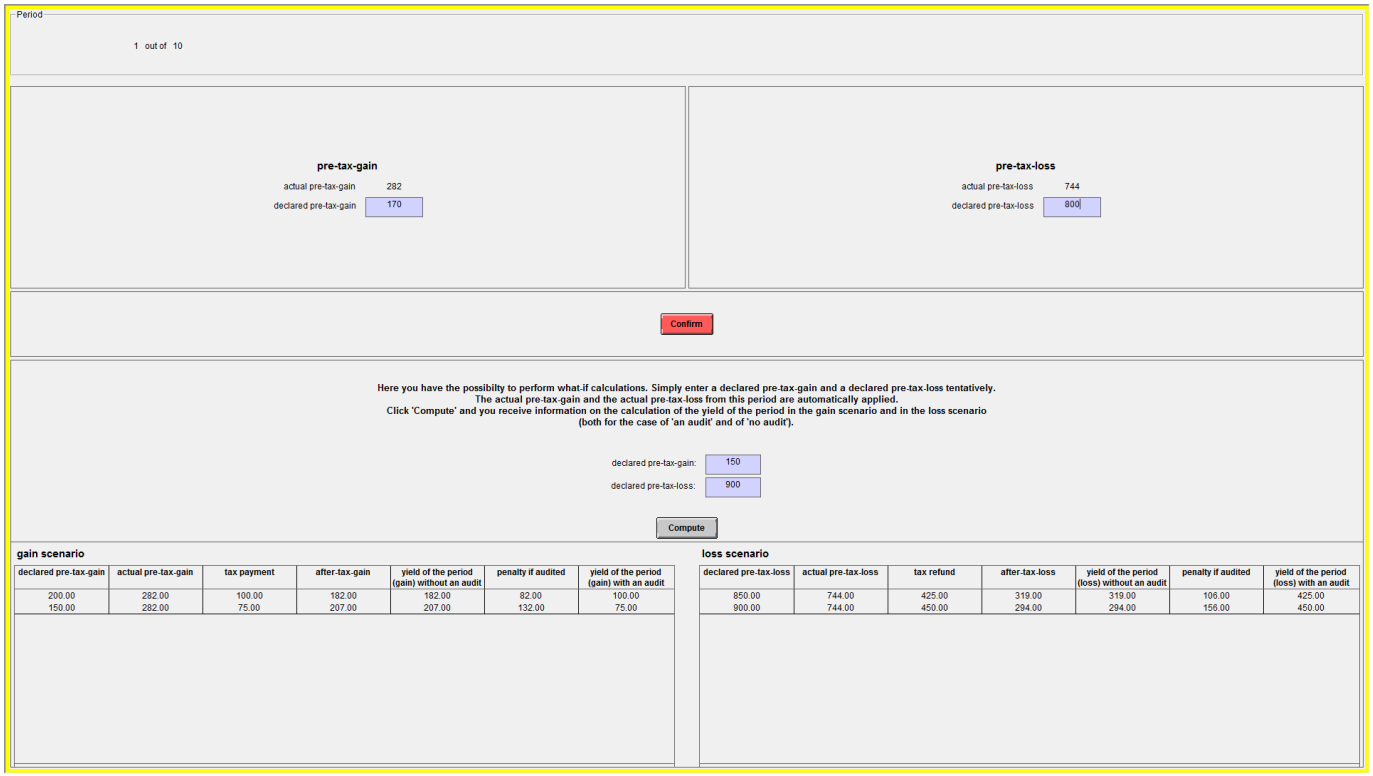
Bevor the real experiment with 10 periods starts there is a rehearsal with 2 training periods. The decisions you make in these training periods have no influence on the payoff of the experiment.

4.9 Appendix B: Screenshots from the Experiment

B1 Screenshots of the Baseline Experiments

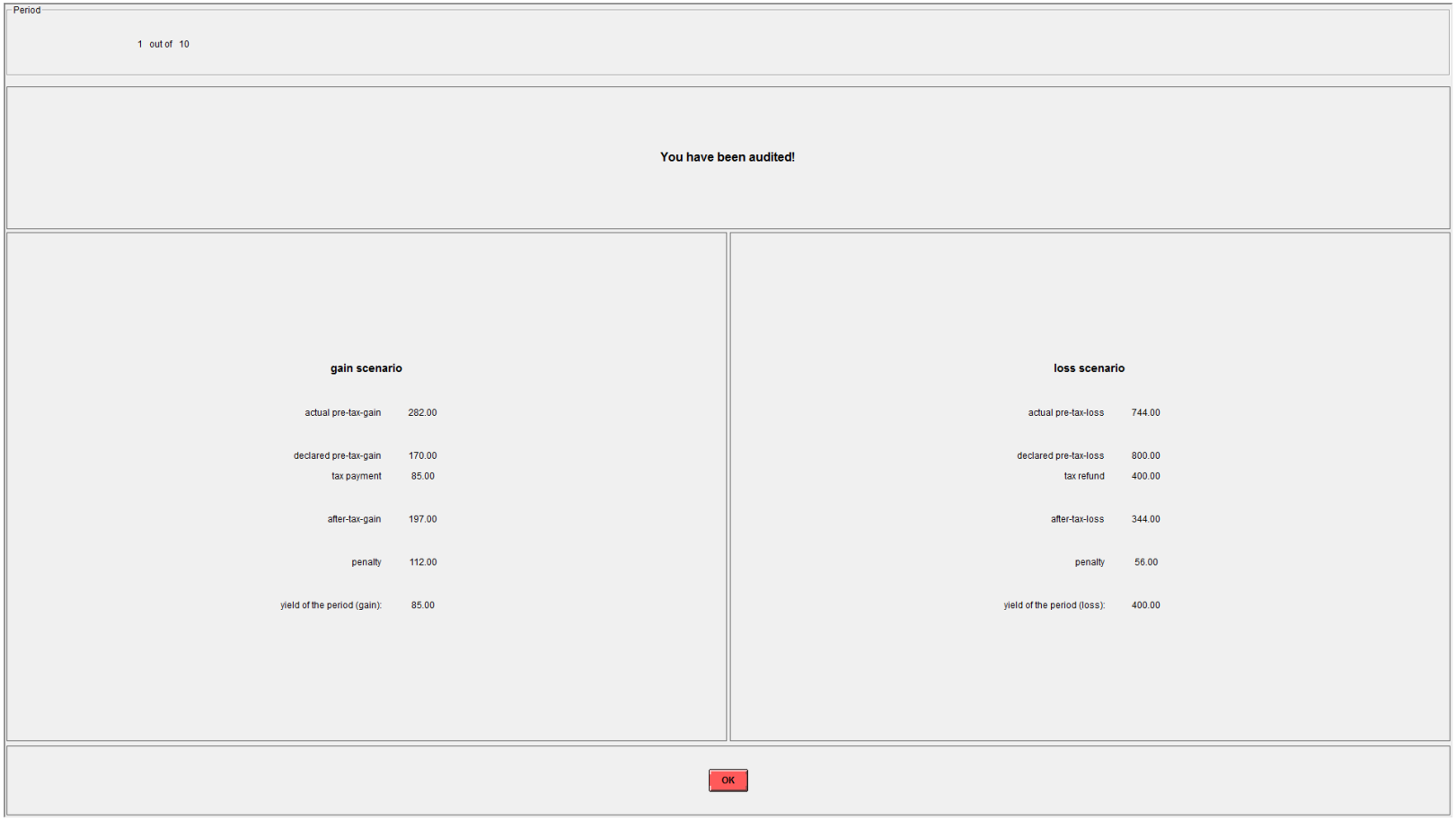
Note that Figure 4.2 and Figure 4.3 display exemplary screenshots for the Baseline ($PI_a \neq |NI_a|$) experiment. However, the only difference to the Baseline ($PI_a = |NI_a|$) experiment is that the two actual pre-tax amounts displayed are identical in the latter experiment.

Figure 4.2: Exemplary Screenshot for the Decision Stage of the Baseline ($PI_a \neq |NI_a|$) Experiment



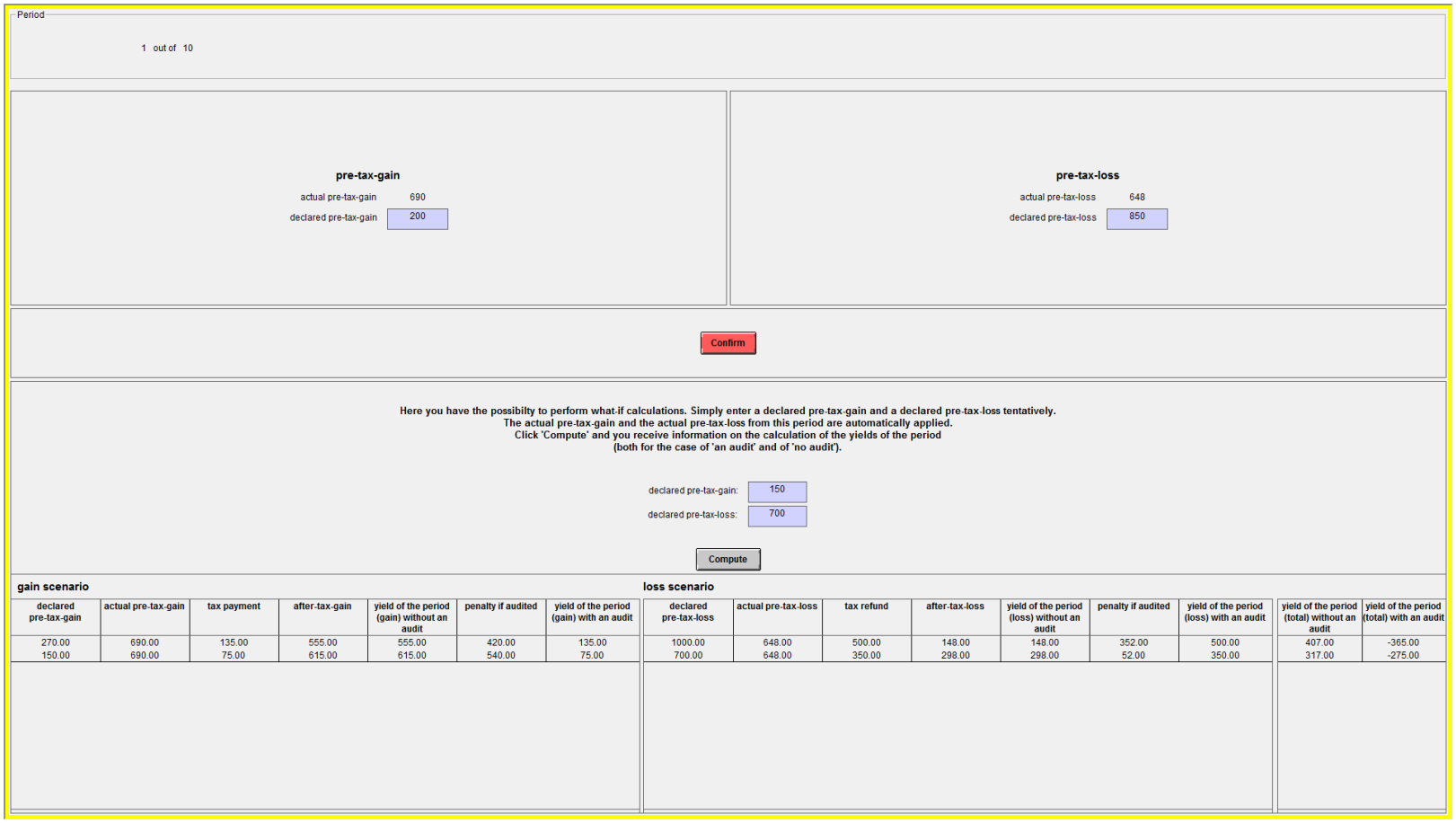
Note: To avoid that participants are overloaded or confused, we decided to fix the income's presentation on the screen in each period (i.e., that the positive income is presented on the left and the negative income on the right hand side). However, we reran the Baseline ($PI_a = |NI_a|$) experiment with a randomized presentation to exclude order effects. Our results are robust to this variation.

Figure 4.3: Exemplary Screenshot for the Information Stage of the Baseline ($PI_a \neq |NI_a|$) Experiment



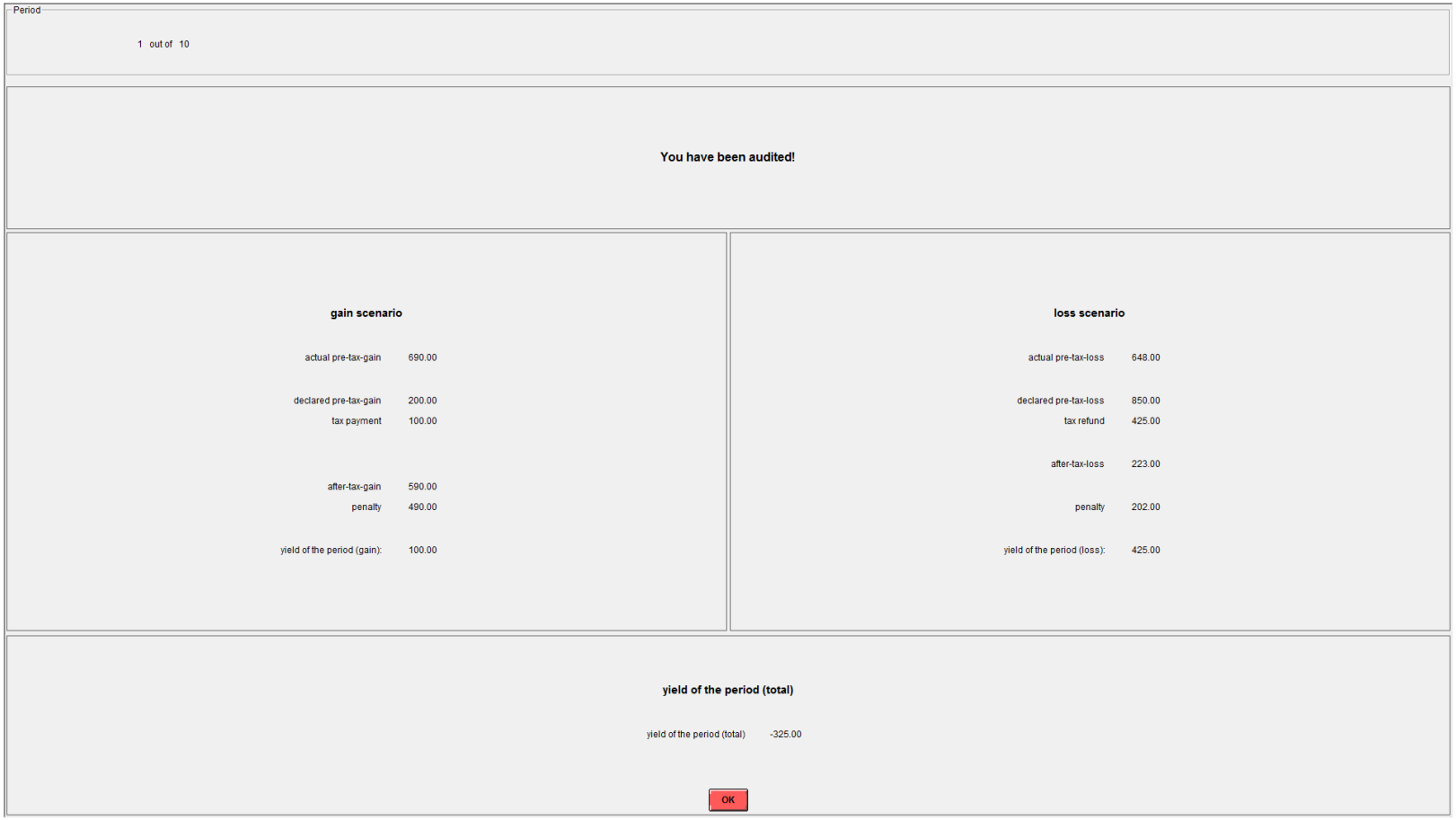
B2 Screenshots of the Offsetting-Low-Salience Treatment

Figure 4.4: Exemplary Screenshot for the Decision Stage of the Offsetting-Low-Salience Treatment



declared pre-tax-gain	actual pre-tax-gain	tax payment	after-tax-gain	yield of the period (gain) without an audit	penalty if audited	yield of the period (gain) with an audit	declared pre-tax-loss	actual pre-tax-loss	tax refund	after-tax-loss	yield of the period (loss) without an audit	penalty if audited	yield of the period (loss) with an audit	yield of the period (total) without an audit	yield of the period (total) with an audit
270.00	690.00	135.00	555.00	555.00	420.00	135.00	1000.00	648.00	500.00	148.00	148.00	352.00	500.00	407.00	-365.00
150.00	690.00	75.00	615.00	615.00	540.00	75.00	700.00	648.00	350.00	298.00	298.00	52.00	350.00	317.00	-275.00

Figure 4.5: Exemplary Screenshot for the Information Stage of the Offsetting-Low-Salience Treatment



B3 Screenshots of the Offsetting-High-Salience Treatment

Figure 4.6: Exemplary Screenshot for the Decision Stage of the Offsetting-High-Salience Treatment

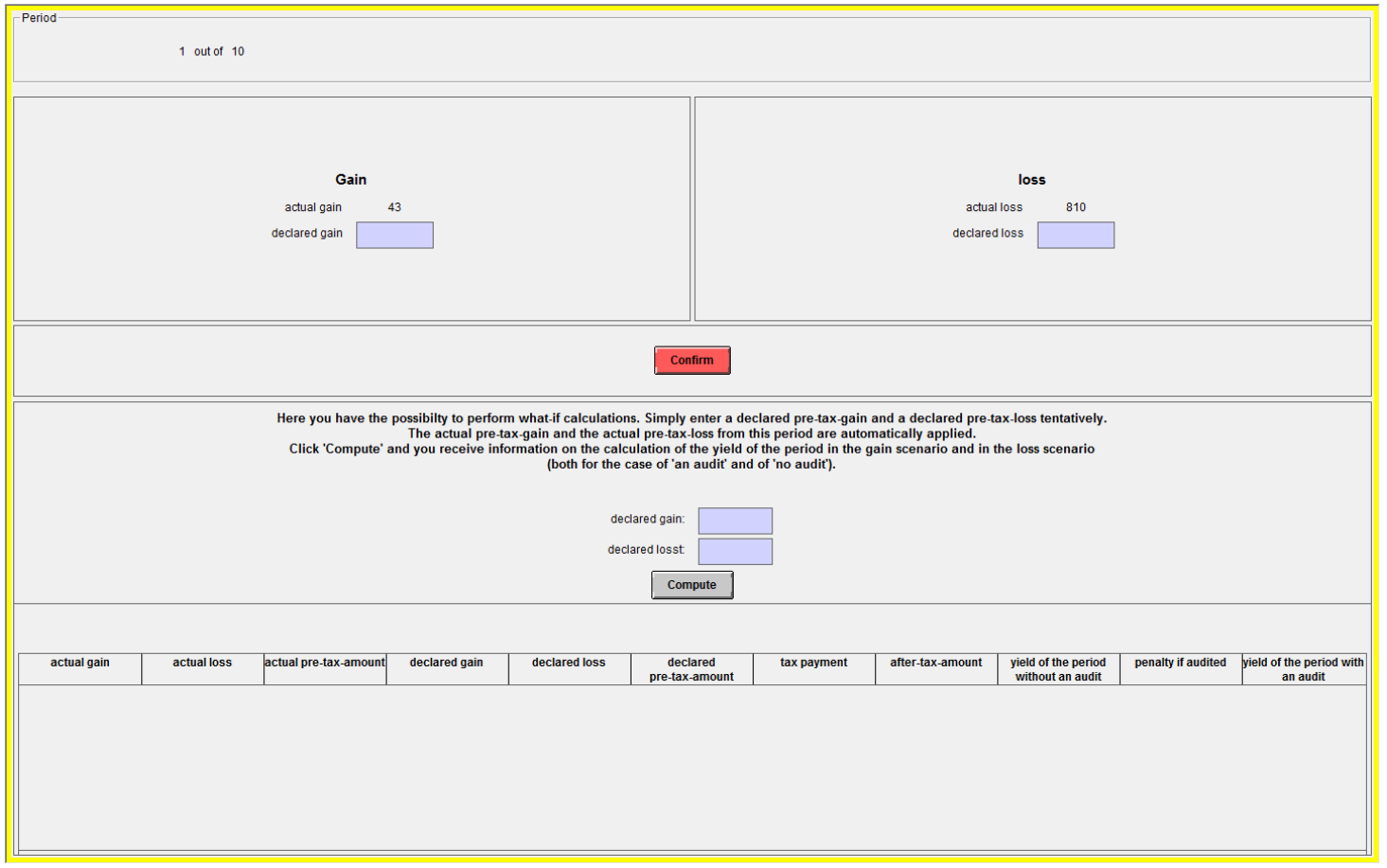
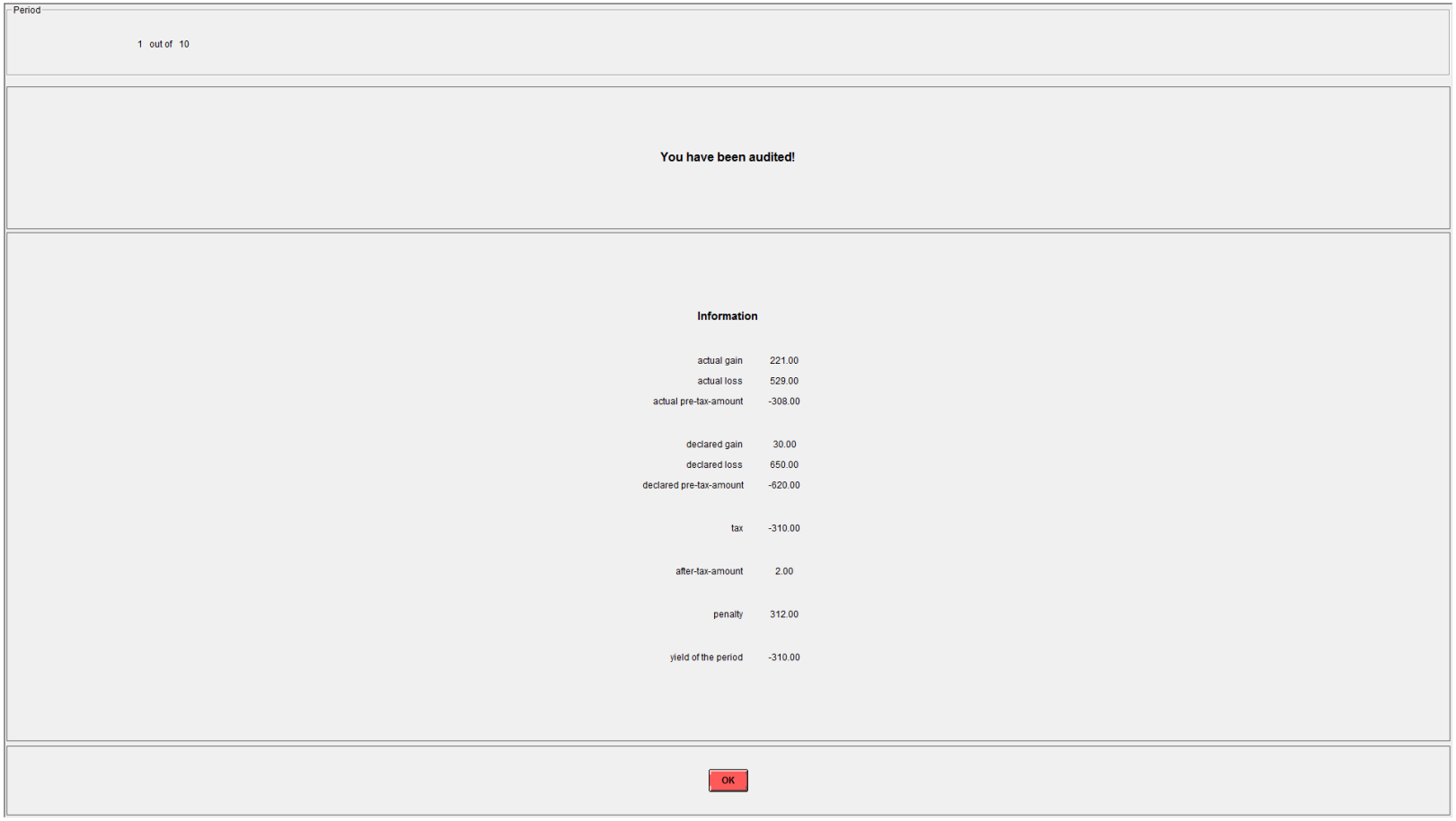


Figure 4.7: Exemplary Screenshot for the Information Stage of the Offsetting-High-Salience Treatment



4.10 Appendix C: Additional Results

Table 4.8: Descriptive Statistics for Individual Characteristics

	experiment 1 (N = 25)	**experiment 2 (N = 34)**	**experiment 3 (N = 146)**
FEMALE	52.0%	29.4%	58.2%
ECONOMICS MAJOR	56.0%	26.5%	42.5%
BACHELOR'S DEGREE	92.0%	82.3%	58.2%
AGE	21.72	23.4	25.9
RISK ATTITUDE	5.6	4.0	5.0
INCOME (IN EURO)	258.2	260.7	369.4
TAX KNOWLEDGE	2.3	2.4	2.7

Note: This table provides an overview of the individual characteristics of our experiments. ECONOMICS MAJOR (BACHELOR'S DEGREE) denotes whether a subject studies economics or management (in a bachelor's degree program). RISK ATTITUDE represents a subject's self-reported willingness to take risk (measured on an 11-point scale, where 0 = not willing to take a risk and 10 = highly willing to take a risk). INCOME is the monthly income after fixed cost. TAX KNOWLEDGE indicates the individual's self-reported proficiency concerning taxes (measured on a 7-point scale where 1 = no knowledge and 7 = wide knowledge).

Table 4.9: Treatment Comparisons – OLS Regressions for our Experiment 3

Dependent Variable	Ratio Of Tax Evasion Difference		Ratio Of Tax Evasion For Positive Income		Ratio Of Tax Evasion For Negative Income		Ratio Of Total Tax Evasion	
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**	**model 7**	**model 8**
Offsetting Treatment	-0.10***		-0.12**		-0.02		-0.07	
	(0.03)		(0.06)		(0.05)		(0.05)	
Offsetting-Low-Salience Treatment		-0.09***		-0.07		0.02		-0.03
		(0.03)		(0.06)		(0.06)		(0.06)
Offsetting-High-Salience Treatment		-0.11***		-0.15**		-0.04		-0.10
		(0.03)		(0.06)		(0.06)		(0.06)
LAST PERIOD	-0.02	-0.02	-0.00	-0.00	0.02	0.01	0.01	0.01
AUDIT	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
PERIOD	0.00	0.00	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ACTUAL POS.	0.04**	0.04**	-0.04	-0.03	-0.07***	-0.07***	-0.07***	-0.07***
INCOME<|ACTUAL	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
NEG. INCOME|								
WHAT IF	0.02***	0.02***	0.02***	0.02***	0.00	0.00	0.01*	0.01*
CALCULATIONS	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
FEMALE	-0.00	-0.00	-0.06	-0.06	-0.06	-0.06	-0.07	-0.07
	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
ECONOMICS	-0.01	-0.01	-0.07	-0.07	-0.05	-0.06	-0.06	-0.06
MAJOR	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
BACHELOR	-0.03	-0.04	-0.11**	-0.12**	-0.08	-0.09*	-0.10*	-0.11**
	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
AGE	-0.00	-0.00	-0.01**	-0.01**	-0.01	-0.01	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK ATTITUDE	0.01	0.01	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
INCOME	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
TAX KNOWLEDGE	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
CONSTANT	0.09	0.09	0.42***	0.44***	0.33**	0.35**	0.39***	0.41***
	(0.07)	(0.07)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
No. of Observations	1,409	1,409	1,410	1,410	1,409	1,409	1,410	1,410
No. of Clusters/ Subjects	141	141	141	141	141	141	141	141
R-squared	0.043	0.043	0.193	0.198	0.164	0.167	0.203	0.207

Note: This table shows the results of OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 4.10: Treatment Comparisons – OLS Regressions for our Experiment 3 (Participant-Averages)

Dependent Variable	**Ratio Of Tax Evasion Difference**		**Ratio Of Tax Evasion For Positive Income**		**Ratio Of Tax Evasion For Negative Income**		**Ratio Of Total Tax Evasion**	
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**	**model 7**	**model 8**
Offsetting Treatment	-0.09***		-0.11*		-0.02		-0.07	
	(0.03)		(0.06)		(0.05)		(0.06)	
Offsetting-Low-Salience Treatment		-0.08**		-0.06		0.02		-0.02
		(0.04)		(0.07)		(0.07)		(0.07)
Offsetting-High-Salience Treatment		-0.10***		-0.14**		-0.04		-0.09
		(0.03)		(0.06)		(0.06)		(0.06)
FEMALE	0.00	0.00	-0.05	-0.05	-0.05	-0.05	-0.06	-0.06
	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
ECONOMICS	-0.02	-0.02	-0.07	-0.07	-0.05	-0.05	-0.06	-0.07
MAJOR	(0.03)	(0.03)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
BACHELOR	-0.04	-0.04	-0.12**	-0.13**	-0.08	-0.09*	-0.10*	-0.11**
	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
AGE	-0.00	-0.00	-0.01**	-0.01**	-0.01*	-0.01*	-0.01**	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK ATTITUDE	0.01	0.01	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
INCOME	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
TAX KNOWLEDGE	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
CONSTANT	0.18**	0.19**	0.54***	0.56***	0.35***	0.37***	0.45***	0.46***
	(0.07)	(0.08)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)
No. of Observations	141	141	141	141	141	141	141	141
R-squared	0.080	0.083	0.284	0.293	0.281	0.287	0.301	0.309

Note: This table shows the results of OLS regressions with standard errors (presented in parentheses). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 4.11: OLS Regressions for our Robustness Experiments (Dependent Variable: RATIO OF TAX EVASION)

	Fear of Further Losses	**Rewording: Revenue and Expenses**	
	No Audit Experiment	**Baseline ($PI_a = \|NI_a\|$)**	**Offsetting-High-Salience**
	model 1	**model 2**	**model 3**
NEGATIVE INCOME	**-0.12*** **	**-0.08** **	**0.01**
	(0.03)	**(0.03)**	**(0.02)**
LAST PERIOD AUDIT	0.13**	-0.03	0.10**
	(0.05)	(0.05)	(0.04)
PERIOD	0.01	0.02**	0.01
	(0.01)	(0.01)	(0.01)
ACTUAL POSITIVE INCOME<|ACTUAL NEGATIVE INCOME| ($PI_a < |NI_a|$)	---	-0.07	0.06
		(0.04)	(0.06)
WHAT IF CALCULATIONS	0.05	0.02	0.01
	(0.04)	(0.02)	(0.02)
FEMALE	0.21	0.18	-0.28***
	(0.15)	(0.12)	(0.09)
ECONOMICS MAJOR	-0.11	-0.08	0.16
	(0.25)	(0.10)	(0.12)
BACHELOR	0.19	-0.21**	-0.12
	(0.16)	(0.09)	(0.07)
AGE	-0.03	0.02*	-0.02***
	(0.03)	(0.01)	(0.01)
RISK ATTITUDE	-0.04	0.08***	0.03
	(0.03)	(0.01)	(0.02)
INCOME	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
TAX KNOWLEDGE	-0.02	-0.00	-0.03
	(0.06)	(0.03)	(0.03)
CONSTANT	1.56**	-0.34	0.56**
	(0.76)	(0.29)	(0.21)
No. of Observations	340	579	439
No. of Clusters/Subjects	17	29	22
R-squared	0.323	0.360	0.266

Note: In this table, the results of linear regressions are presented for our robustness experiments with RATIO OF TAX EVASION as dependent variable. To take into account that subjects face repeated decision situations, we run OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 4.12: OLS Regressions for our Robustness Test "Bunching" (Dependent Variable: RATIO OF TAX EVASION)

Description	**Reported negative income of 1,000 is excluded**			**Reported negative income between 900 and 1,000 is excluded**		
Experiment	**1**	**2**	**3**	**1**	**2**	**3**
	model 1	model 2	model 3	model 4	model 5	model 6
NEGATIVE INCOME	**-0.13****	**-0.06****	**-0.04*****	**-0.10***	**-0.06****	**-0.03****
	(0.05)	**(0.03)**	**(0.01)**	**(0.05)**	**(0.03)**	**(0.01)**
LAST PERIOD AUDIT	0.09	0.09**	0.01	0.06	0.12**	0.01
	(0.05)	(0.04)	(0.02)	(0.05)	(0.05)	(0.02)
PERIOD	-0.00	0.03	0.01*	-0.00	0.03	0.02**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
ACTUAL POSITIVE INCOME < |ACTUAL NEGATIVE INCOME| ($PI_a < \|NI_a\|$)	-0.00	0.01**	0.01***	-0.00	0.00	0.01**
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
WHAT IF CALCULATIONS	---	-0.05	-0.05**	---	-0.00	-0.02
		(0.04)	(0.02)		(0.04)	(0.03)
FEMALE	-0.19*	-0.13*	-0.07	-0.19*	-0.15*	-0.07
	(0.10)	(0.08)	(0.05)	(0.10)	(0.08)	(0.06)
ECONOMICS MAJOR	-0.09	0.05	-0.05	-0.11	0.04	-0.05
	(0.13)	(0.08)	(0.06)	(0.14)	(0.08)	(0.06)
BACHELOR	-0.55***	0.10	-0.09	-0.56***	0.10	-0.10*
	(0.12)	(0.09)	(0.05)	(0.13)	(0.09)	(0.06)
AGE	-0.04	0.01	-0.01*	-0.04	0.01	-0.01
	(0.02)	(0.01)	(0.00)	(0.02)	(0.02)	(0.00)
RISK ATTITUDE	0.04	0.02	0.06***	0.04	0.02	0.06***
	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)
INCOME	0.00*	-0.00	0.00	0.00**	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
TAX KNOWLEDGE	-0.00	0.00	0.00	0.01	0.01	0.01
	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.02)
CONSTANT	1.60**	-0.11	0.39***	1.67**	-0.05	0.38***
	(0.67)	(0.36)	(0.13)	(0.71)	(0.38)	(0.14)
No. of Observations	482	650	2,819	430	585	2,443
No. of Clusters/Subjects	25	34	141	25	34	141
R-squared	0.229	0.125	0.173	0.241	0.132	0.176

Note: In this table, the results of linear regressions are presented with RATIO OF TAX EVASION as dependent variable. To take into account that subjects face repeated decision situations, we run OLS regressions with robust standard errors (presented in parentheses) clustered at the subject level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Chapter 5

The Effect of Self Disclosure Opportunities on Tax Compliance – An Experimental Investigation*

Abstract

This paper experimentally examines the influence of the self-disclosure opportunity on tax evasion. Objectors of a self-disclosure opportunity argue that this increases tax evasion as it reduces tax morale and is used strategically. Using a two-stage tax evasion game, we do not find any crowding out effect of formerly honest tax payers, but find that subjects strategically integrate the self-disclosure option into their tax evasion decision if audit probabilities are uncertain, thus increasing tax evasion. However, we show that small penalty payments that accompany the self-disclosure opportunity are sufficient to limit this tax evasion increase. Considering that the self-disclosure option can be a useful tool to offer subjects a way back to tax honesty and can generate revenues that governments would not have been able to generate within their standard audit capacities, the self-disclosure opportunity that includes small penalty payments provides an instrument for governments to increase tax revenues without bearing a considerable fiscal risk.

Keywords: Voluntary Self-Disclosure Tax compliance · Tax evasion · Experimental Economics

JEL Codes: C91 · D91 · H20 · H24 · H26 · H30 · K42

* This chapter is single-authored.
We thank Kay Blaufus, Neela Cramer, Martin Fochmann, Jochen Hundsdoerfer, Alexander Schwäbe and Caren Sureth-Sloane, as well as participants of the 2016 arqus Conference.

5.1 Introduction

This paper experimentally examines the influence of the self-disclosure opportunity on tax evasion. The option of voluntary self-disclosure was introduced for mainly two reasons fighting tax evasion. On the one hand, governments want to raise tax revenues in the short run as they expect corrections which they were not able to generate within their standard audit capacities. The administrative costs when inspecting information given in a self-disclosure are substantially lower than when determining taxes based on whistle-blowing information (Langenmayr, 2017). On the other hand, tax amnesties and voluntary disclosure options offer tax evaders who regret their misdeed a gentle way back to honesty, and the government might be able to underpin that tax evasion is a criminal act which indirectly affects every citizen if missing contributions lead to reduced public investments (Parle and Hirlinger, 1986; Hasseldine, 1998b; Alm, 2012).

Against this initial idea of increasing tax revenues, the drawbacks of the self-disclosure opportunities are often raised. First, it is said to erode the tax morale of subjects who would be tax compliant without these opportunities and cause a crowding out of formerly intrinsic motivated tax payers (see e.g., Feld and Frey, 2002, 2007; Vihanto, 2003). Second, it can be strategically used if subjects wait for information on high audit probabilities if e.g., tax data carriers are bought, and disclose their evaded taxes only under these circumstances. Both arguments lead to an increase of tax evasion. Therefore, this study aims at analyzing whether the criticism of voluntary self-disclosure is justified as it indeed decreases tax compliance and whether a small penalty which accompanies the self-disclosure is sufficient to limit the probable negative effects.

In order to analyze both points of criticism separately, we have designed a two-stage tax evasion game with repetition and implemented three treatments. The first treatment does not offer any voluntary disclosure, the second one does offer this opportunity and is granting impunity, the third one offers a self-disclosure opportunity but levies a small penalty payment on the disclosed taxes. In the first part of the experiment, subjects have to make tax evasion decisions under certainty, i.e., they have full information on all variables before making their decision and all variables (including the audit probability) are fixed. As under these circumstances, the self-disclosure opportunity may not be used strategically, we may control for an erosion of tax morale in this part if we find a different tax evasion behavior. In the second part of the experiment, participants still have full information on the decision variables but we integrate uncertainty regarding the actual audit probability. Here, before making their tax evasion decision subjects are presented two audit probabilities which both might occur with the same probability. Only after their tax evasion decision, subjects are informed about the actual drawn audit probability and may revise their decision in the corresponding treatments. By integrating uncertainty, we are able to analyze whether subjects indeed use the self-disclosure opportunity strategically and whether this aspect leads to decreased tax compliance. By integrating the third treatment which offers a self-disclosure option while levying a small penalty on the evaded taxes we test whether small penalties

solve the described potential problems of an eroded tax morale and strategic use of the self-disclosure opportunity.

To our knowledge, the only paper who examines voluntary disclosure effects on tax evasion is Langenmayr (2017) who analyzes data of the 2009 offshore voluntary disclosure and confirms that it reduced tax compliance while she finds also evidence that administrative costs are reduced. By nature, data on tax evasion is very limited and tax evasion therefore has to be measured by using proxies such as the taxable income or deposits of citizens of offshore bank accounts. In order to overcome these limitations, we use a controlled experimental environment in order to be able to measure tax evasion properly as it is completely observable there. Moreover, we are able to manipulate the subject's environment (i.e., whether she is offered a voluntary disclosure opportunity or not) while keeping all other external parameters (e.g. audit rates, penalty levels, tax rate etc.) equal for all subjects. Above, we are able to introduce an environment with certain and uncertain audit probabilities to analyze the impact of intrinsic morale erosion and of the strategic use of the self-disclosure option on tax compliance.

Against our hypothesis, we do not find an erosion of tax morale which reduces tax compliance if subjects are offered a self-disclosure opportunity. However, if subjects can use the self-disclosure opportunity strategically as they face uncertain audit probabilities tax compliance decreases. In addition, we find that if the self-disclosure opportunity is not offered with impunity, but is accompanied by a rather small penalty payment, subjects increase tax evasion only moderately, i.e., within the expected range. Thus, confirmed by our additional analyses we may conclude that the omission of a penalty on the evaded taxes leads to a legalization of tax evasion itself. Correspondingly, even a small penalty payment that accompanies the self-disclosure opportunity signalizes the prohibition of tax evasion and may be able to keep tax evasion in calculable bounders.

The remaining part of this paper is organized as follows. In the next section, we give an overview of prior studies on the impact of tax amnesty and voluntary disclosure on tax evasion. In the third section, we outline our experimental design, derive the hypotheses, present the experimental protocol and sample characteristics and introduce the variable measurement. Section four presents the descriptive statistics and results for multivariate analyses including some robustness checks. In the fifth section, we run several additional analyses to investigate the possible factors that drive the outcomes of the hypotheses' analyses. The sixth section concludes the paper.

5.2 Tax Amnesty and Voluntary Disclosure Research

So far, the vast majority of literature dealing with subsequent declaration of withheld taxes analyzes tax amnesties which occurred once or in a series. However, the literature on voluntary disclosures is very limited. Based on the analytical seminal paper of Allingham and Sandmo (1972), who suppose that rational individuals maximize their expected utility after taxes by taking the given audit probabilities and penalty rates into account, some papers investigate the influence of tax amnesties on tax evasion behavior. Malik and Schwab (1991) integrate the probability of a tax amnesty into the economic model of Allingham and Sandmo (1972) as well as two different ex ante utility functions. Based on these utility functions, the subject has to make the tax evasion decision. Afterwards, the subject is presented the ex post utility function whereby she either faces small (fiscal) disadvantages or larger costs, e.g. morale constraints that hamper the tax evasion advantage. Based on the ex post utility function the subject is granted the amnesty with a certain probability.[1] If the subject chooses the amnesty when offered, she declares her whole income truthfully. The authors find that tax evasion increases with the probability of a tax amnesty. Furthermore, they integrate uncertainty about the audit level by applying ex ante a high and a low audit probability in order to control for a strategic use of the amnesty. Only after subjects have made their tax evasion decision, they are informed about the actual audit probability. Malik and Schwab (1991) find that subjects are more tax compliant if they face a high audit probability with certainty.

Stella (1991) examines the impact of an amnesty on the tax revenues and finds that an amnesty will most likely reduce revenues unless the state convincingly raises audit probabilities afterwards. This finding is supported by Alm and Beck (1990) who additionally emphasizes that tax amnesty may increase future tax revenues if the taxpayer adopts paying taxes as social norm. However, the amnesty may also reduce tax compliance if the taxpayer anticipates future amnesties with limp tax enforcement. Andreoni (1991) examines the effect of a permanent amnesty, which can be seen as self-disclosure opportunity, on tax compliance and tax revenues. He states that although tax evasion rises due to the permanent amnesty the impact on the tax revenues depends on the pre-amnesty evasion level. If this level is large, the amnesty generates those missing tax revenues and might be profitable.

Besides these theoretical studies, there are some studies who investigate the influence of tax amnesties on tax compliance experimentally. In a public good experiment with seven different treatments, Alm et al. (1990) grant subjects an amnesty in four of these treatments whereby a possible amnesty was announced beforehand only in two treatments. They also vary the level of post-amnesty tax enforcement and whether they grant information on the existence of only a one-time amnesty. As already predicted by theory, they find that a tax amnesty lowers tax compliance and that this decline is mostly driven by taxpayers who were moderate compliant before the amnesty and reduce their tax payments afterwards accordingly. They furthermore find evidence for an anticipation effect which reduces tax compliance before an amnesty if it is expected to occur and that increased tax enforcement after the amnesty leads

[1] In contrast to the voluntary self-disclosure opportunity, an amnesty only occurs with a certain probability.

to an increase of tax compliance in the aftermath.[2] Torgler and Schaltegger (2005) adopt the experimental design of Alm et al. (1990) and introduced the opportunity to discuss and vote for or against the amnesty. They run their experiments in Switzerland and Costa Rica and find that subjects are more compliant if they discuss the amnesty option before voting on it as it might raise the awareness of the economic importance of tax revenues. Providing the voting option without the discussion opportunity offers ambiguous results. However, in contrast to Alm et al. (1990), they do not find that increased tax enforcement after the amnesty increases tax compliance.

Rechberger et al. (2010) examine the effect of tax amnesty's justice perception on tax compliance. They conduct a standard tax compliance game consisting of ten periods and an amnesty after the fifth period whereby the subjects are held uninformed about the number of periods and the existence of the amnesty opportunity. After the amnesty is executed subjects had to state their justice perception and play another five periods without enhanced tax enforcement. In contrast to all other presented studies, Rechberger et al. (2010) find that tax compliance increases after the tax amnesty although the experimenters do not increase tax enforcement. They additionally find that the perceived amnesty's justice positively influences the post-amnesty's tax compliance as well as the retribution and value restoration.[3]

The most recent and as far as we are aware first paper on voluntary self-disclosure is developed by Langenmayr (2017) who first examines self-disclosure mechanisms theoretically and analyses her results empirically afterwards. The author models a framework which includes a decision diagram considering the government's tax parameters, the subject's decisions and nature's draws as well as subject's morale costs. In line with the theoretical papers on tax amnesties she finds that the existence of a voluntary self-disclosure increases tax evasion but may increase tax revenues of net administrative costs if the government increases detection probability and fine simultaneously. Langenmayr (2017) confirms the relevance of administrative costs by carrying out a survey among all regional tax officers in Germany. About 60% of the tax officers state that administrative effort is significantly lower after a voluntary disclosure in order to assess previously evaded taxes. In a last step she empirically analyzes

[2] The positive effect of tax enforcement after an amnesty on tax compliance was also shown in empirical analyses of Alm and Beck (1991), Christian et al. (2002) and Battiston et al. (2020) but relativized by Fisher et al. (1989), Alm and Beck (1993) and López-Laborda and Rodrigo (2003) who do not find any short- or long-term effects of an amnesty on tax compliance as a stronger post-amnesty tax enforcement is offset by the general compliance reducing effect of the amnesty. Furthermore, Luitel and Sobel (2007) find that repeated tax amnesties also generate less short-term revenue as they increase revenue losses due to disincentives for long-term tax compliance. Andersson et al. (2019) study the impact of tax information exchange agreements between Norway and tax heavens on amnesty applications. They find that the use of voluntary disclosure increases when stricter tax enforcements are announced even if the execution takes place much later.

[3] All experimental papers presented have in common that if an audit occurs or an amnesty is granted more than one period is affected. That means, if the subject is audited in one period the last five (in Rechberger et al., 2010 only three) periods are checked and the penalty is levied on all these periods. If a subject uses the amnesty, she has to pay the missing taxes for the last five (three) periods. Thus, periods are not independent and the tax evasion decision in one period is always based on the last periods' decisions and the audit and amnesty expectation of the future periods. However, none of these papers has taken this interdependency into account accordingly when analyzing the data.

data of the 2009 offshore voluntary disclosure program in the U.S. and confirms that it reduced tax compliance.

5.3 Hypotheses Development, Experimental Design and Protocol, and Variable Measurement

5.3.1 Experimental Design and Hypotheses Development

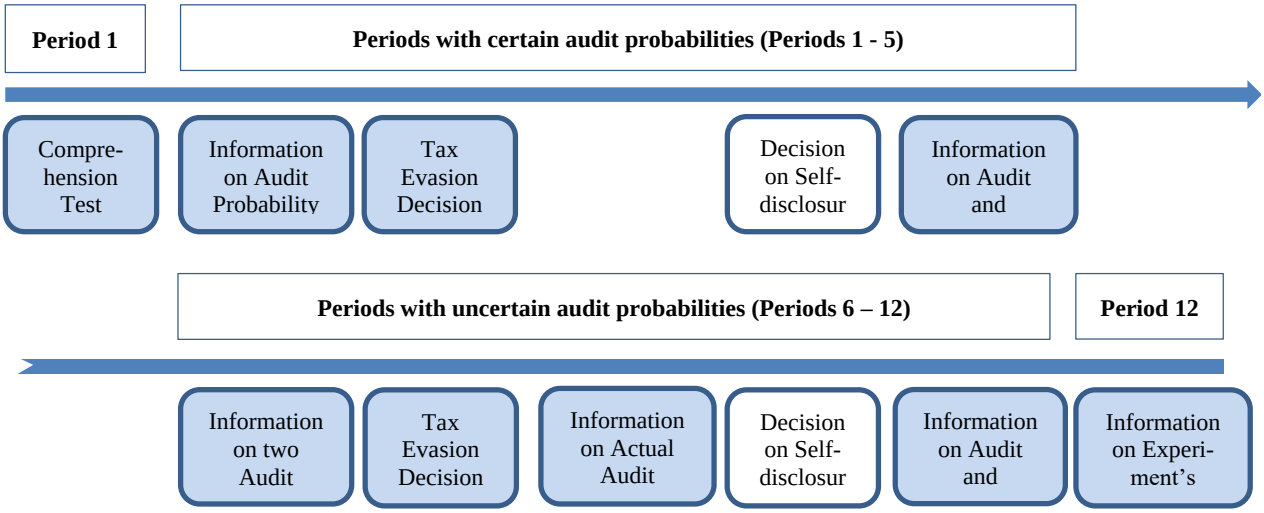
We implement a tax evasion game in which subjects are randomly assigned to either the standard game without any voluntary self-disclosure opportunity or a tax evasion game with a voluntary disclosure opportunity (between-subject design).[4] Each treatment is first run with five periods containing certain audit probabilities in order to test whether the intrinsic motivation to pay taxes is undermined by the self-disclosure option. These certain periods are followed by seven periods with uncertain audit probabilities in which we test whether subjects use the option strategically and increase tax evasion (within-subject variation). After being assigned to personal computers, subjects receive written instructions on the experiment.

Before the participants start the experiment they have to answer several comprehension questions to ensure that they fully understand the experiment. After each tax evasion decision subjects are informed about whether they were audited, the respective penalty payments and the resulting period's payoff. At the very end of the twelfth period, the computer randomly draws one period which is relevant for payoff. Afterwards, subjects are asked to complete a questionnaire that collects socio-demographic data.[5] Additionally, they earn a show-up fee of 3 euros. The detailed experiment's procedure is illustrated in Figure 5.1.

[4] The experimental software was programmed with z-Tree (Fischbacher, 2007).

[5] The experiment's instructions are translated and presented in Section 5.7 (Appendix A), exemplary screenshots are displayed in Section 5.8 (Appendix B) and the respective questionnaire is presented in Section 5.9 (Appendix C).

Figure 5.1: Experiment's Overview



Note: The decision on self-disclosure (white box) is not granted for subjects in the No Self-Disclosure treatment.

In each period subjects receive an initial endowment E of 1,000 cents (10 euros). Implementing a tax rate of 40% subjects should pay a tax T of 400 cents.[6] Even so, subjects are free to actually pay any integer amount of taxes D between 0 and 400 cents. In the first five periods, subjects are audited with a certain audit probability p that is presented before the subject makes her tax evasion decision and that fluctuates between 0% and 80%.[7] If a subject is caught cheating she has to pay the evaded tax and a penalty F which amounts to 100% of the evaded taxes. If the subject declared the income truthfully, no additional payments are levied. Thus, the subject's expected payoff in the No Self-Disclosure Treatment under certainty results in

$$\mathrm{E}[\pi_{NoSD}^{Certain}] = (1 - \mathrm{p})(\mathrm{E} - \mathrm{D}) + \mathrm{p}[\mathrm{E} - \mathrm{D} - (T - D)(1 + F)] \tag{5.1}$$

whereby the payoff if no audit takes plays is displayed by

$$\pi_{NA} = \mathrm{E} - \mathrm{D} \tag{5.2}$$

and the payoff if an audit takes plays results in

$$\pi_{A} = \mathrm{E} - \mathrm{D} - (T - D)(1 + F). \tag{5.3}$$

If the subject is fully compliant, her payoff is presented by

$$\pi_{C} = \mathrm{E} - \mathrm{T}. \tag{5.4}$$

[6] The taxes levied (including potential penalties) are used for further research at the Leibniz University of Hannover.

[7] Please note, that we use neutral language in our experiment to diminish the impact of loaded terms on subjects' decisions (see e.g., Alm, 1991, 2010). Hence, "taxes" are labelled "contribution", "audit" is called "control" etc.

All treatments' decisions and the respective (expected) payoffs are presented in Table 5.4. Assuming risk-neutral and income maximizing subjects we compare the payoff in case of tax compliance (equation 5.4) with the expected payoff in case of tax evasion (equation 5.1) to derive the critical audit probability under which tax evasion is no longer optimal in the No Self-Disclosure Treatment under certainty:[8]

$$\mathrm{p}^* = \frac{1}{1+F}. \tag{5.5}$$

As the penalty F amounts to 100% in our experimental setting, complete tax evasion is optimal as long as the audit probability is below 50%, whereas tax compliance is optimal for all audit probabilities above 50%.

Subjects who are assigned to the voluntary disclosure treatments additionally have the opportunity to revise their tax evasion decision and declare their income truthfully after their initial declaration, but before the audit takes place. If the subject uses the self-disclosure option in the Self-Disclosure without Penalty Treatment, she only has to repay the evaded tax but no additional penalty is levied (f = 0%). If the subject uses the self-disclosure option in the Self-Disclosure with Penalty Treatment, she has to repay the evaded tax but only has to pay a penalty f that amounts to 10% of the evaded tax (instead of 100%). The respective payoff for the self-disclosure usage results in:

$$\pi_{SD} = \mathrm{E} - \mathrm{D} - (T - D)(1 + f) = E - T - (T - D)f. \tag{5.6}$$

As we grant impunity in the Self-Disclosure without Penalty Treatment, the payoff under self-disclosure (equation 5.6, f = 0%) equals the payoff under full compliance (equation 5.4). However, as we keep all parameters constant within a decision period under certainty, we should not expect subjects to revise their former tax evasion decision and use the self-disclosure opportunity (even if we grant impunity) as long as the subjects do not face an altered risk-perception when making their self-disclosure decision. This approach was analyzed by Malik and Schwab (1991) who introduce two different possible utility functions which each occur with a certain probability. They motivate this step by either different absolute risk aversion levels or an ex ante unknown response of others to tax evasion which changes the subject's utility function ex post. In contrast, in the Self-Disclosure with Penalty Treatment (f = 10%) the payoff in the case of ex ante full compliance exceeds the payoff in the case of a voluntary self-disclosure for any $T > D$ (ex post full compliance). Hence, it is not rational to use the voluntary disclosure opportunity if all parameters, such as the audit probability, are certain.[9] Instead, we expect risk-neutral and income maximizing subjects to determine their tax evasion decision on the certain parameters and do not revise their decision in the aftermath. This is also displayed in Table 5.4 when comparing the first two treatment variations. As the self-disclosure option cannot be used strategically in periods with certain audit probabilities, we can conclude a repressed tax morale, if we find that

[8] For a detailed calculation see Section 5.10 (Appendix D, derivation 1).

[9] As the self-disclosure option in the Self-Disclosure treatments under certainty is not relevant for risk-neutral and income maximizing subjects, all treatments only have the real alternatives of full compliance (equation 5.4) and full tax evasion (equation 5.1) which consequently lead to a critical audit probability of p^*=50%.

subject's tax compliance is lower in the Self-Disclosure Treatments than in the No Self-Disclosure Treatment. This can be the case, if actually honest taxpayers perceive the self-disclosure option as unfair or it signalizes a legalization of tax evasion itself, so that their intrinsic motivation to pay taxes is crowded out.[10] Thus, we formulate our first hypothesis regarding certain audit probabilities accordingly:

Hypothesis 1: *The voluntary self-disclosure opportunity leads to an erosion of tax morale which decreases tax compliance.*

In a second step we extent the experiment and integrate uncertain audit probabilities in order to test for the strategical use of the self-disclosure opportunity. Whereas in the first five periods audit probabilities are certain, subjects face uncertainty in the last seven periods (within-subject variation) based on Malik and Schwab (1991). In each of these last periods we present two different audit probabilities whereby the high audit probability p_H occurs with a probability $\theta$of 50%. Consequently, the low audit probability p_L occurs with a probability of $1-\theta$ of 50%. Again subjects have to make their tax evasion decision now based on the two probabilities. Afterwards, the computer randomly draws either the high or the low audit probability and presents this result to the subject. For subjects without a self-disclosure option, the expected payoff under uncertain audit probabilities results in

$$\mathrm{E}\left[\pi_{NoSD}^{Uncertain}\right] = \theta[p_H\pi_A + (1-p_H)\pi_{NA}] + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}]. \quad (5.7)$$

Assuming risk-neutrality, subjects in the No Self-Disclosure Treatment should evade all taxes as long as the expected audit probability $\frac{p_H+p_L{}^*}{\theta}$ does not exceed 50%.[11] This result is identical with the optimal tax evasion strategy under certain audit probabilities presented above and displayed in Table 5.4 when comparing the critical audit probabilities (last column) of the first three treatment variations.

Subjects in the Self-Disclosure Treatments subsequently have the opportunity to voluntarily disclose their evaded taxes after being informed about the actual audit probability. In order to determine the optimal tax evasion strategy for the Self-Disclosure Treatments under uncertainty, we use backward induction on the basis of the decision tree presented in the last treatment variation in Table 5.4. Thus, we first analyze when subjects in the Self-Disclosure Treatments under Uncertainty should actually use the self-disclosure option. At that moment when subjects decide on whether to use the self-disclosure option or not, they are already informed about the actual drawn audit probability. Hence, they base their decision on certain audit probabilities. This decision was already presented above and displayed in the second treatment variation in Table 5.4 (Self-Disclosure Treatment under Certainty). Thus, even for the self-disclosure decision in the Self-Disclosure Treatment under Uncertainty, we compare the payoffs

[10] For studies on a possible crowding out effect of intrinsic motivation see e.g., Alm et al. (1990), Gneezy and Rustichini (2000), Fehr and Falk (2002), Falk and Kosfeld (2006), Belot and Schröder (2015), Blaufus et al. (2016) and Dwenger et al. (2016).

[11] For a detailed calculation see Section 5.10 (Appendix D, derivation 2).

after a tax evasion decision with the self-disclosure option's use (equation 5.6) and without this use (equation 5.1). We find that the critical audit probability amounts to 50% for the Self-Disclosure without Penalty Treatment and to 55% for the Self-Disclosure with Penalty Treatment.[12] For audit probabilities above this level, subjects should opt for the voluntary self-disclosure option, for audit probabilities below this level it is optimal not to use the self-disclosure option.

We have modelled the audit probabilities in such a way that full tax evasion without using the self-disclosure opportunity is optimal if p_L is drawn, whereas it is optimal using the opportunity if p_H is drawn. Those two decision paths are highlighted in the decision tree for the last treatment variation in Table 5.4 with red lines. We may now determine the optimal tax evasion strategy for the Self-Disclosure Treatments under uncertain audit probabilities (first decision of the decision tree) by comparing the expected value of tax evasion

$$\mathrm{E}\left[\pi_{SD}^{Uncertain}\right] = \theta\pi_{\mathrm{SD}} + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}] \tag{5.8}$$

with the certain payoff under tax compliance (equation 5.4).[13] Thereby, we assume that for all high audit probabilities (that are drawn with the probability θ) the risk-neutral subjects choses the self-disclosure option, whereas she chooses not to use this option if the low audit probability is drawn. The calculation reveals that for all *low* audit probabilities up to 50% (45%) in the Self-Disclosure without (with) Penalty Treatment complete tax evasion is always optimal. In contrast, subjects in the No Self-Disclosure Treatment should only evade taxes if the *expected* audit probability is below 50%, so that we find a divergent tax evasion strategy for the No Self-Disclosure Treatment and the Self-Disclosure Treatments. Thus, we derive our second hypothesis for uncertain audit probabilities:

> *Hypothesis 2:* *The self-disclosure opportunity is strategically integrated into the tax evasion decision, so that tax compliance decreases under initially uncertain audit probabilities.*

As described above, we test whether we find the often mentioned points of criticism of the self-disclosure opportunity – an erosion of tax morale which leads to a crowding out of actually honest taxpayers and the strategical use of the option – which are both said to increase tax evasion. However, we additional test whether there is a probate tool to counter these effects. Therefore, we designed the third treatment – the Self-Disclosure with Penalty Treatment which is identical to the Self-Disclosure without Penalty Treatment but differs only by levying a small penalty (10%) on the disclosed taxes.[14] There are two approaches why we expect the penalty payment to reduce tax evasion. First, the existence of a possible penalty increases the complexity of the tax evasion decision compared to an environment in which the subject is granted a self- disclosure option with impunity (Dohmen et al., 2010). Numerous

[12] For a detailed calculation see Section 5.10 (Appendix D, derivation 3).
[13] For a detailed calculation see Section 5.10 (Appendix D, derivation 4).
[14] If being detected with tax evasion subjects in all treatment have to pay a penalty of 100% on the evaded taxes.

studies already have shown that complexity (cognitive load) decreases risk-taking.[15] Hence, complexity should also reduce tax evasion. Second, if the existence of a penalty on the self-disclosed evaded taxes makes tax evasion appear more illegal than under the self-disclosure option without penalty, the illegalization of the self-disclosure opportunity can illegalize tax evasion behavior itself, thus increasing tax compliance.[16] Thus, we formulate the third hypotheses:

Hypothesis 3: *Small penalty payments which accompany the self-disclosure option are sufficient to curb tax evasion associated with this option.*

5.3.2 Experimental Protocol and Sample Characteristics

The participants were recruited with the software hroot (Bock et al., 2014) and assigned themselves into one of nine sessions which were conducted at the computerized experimental laboratory of the Leibniz University of Hannover. Participants assign themselves to a workstation by drawing a table tennis ball when entering the room. After they were seated at the respective workstation, general information on basic rules within the experiment were loudly spoken by the experimenter (e.g., that no communication was aloud, questions were only to be asked and answered with the experimenter in private by raising the hand etc.). Afterwards, written instructions were handed out, which explain the experiment's procedure, the decision task, the payout rule and give information on the comprehension question and questionnaire. After reading the instructions carefully, participants are offered to ask questions on the experiment in private and are told that they have the opportunity to do so at any time during the experiment.

During the experiment parameters such as the initial endowment E, the fine rate F (and f) and also the allocation to the respective treatment were fixed. However, we varied the audit rate over the periods to set incentives for varied behavior and to test for our hypotheses. The audit probabilities per period as well as the respective expected tax evasion rate and self-disclosing strategy (if allocated to one of the two respective treatments) assuming risk-neutral and income maximizing individuals is presented in Table 5.1. In the first five periods, we alter the certain audit probabilities between 0% and 80% in 20% increments. As shown above, the critical audit probability is 50% for all treatments, so that full tax evasion is optimal for audit probabilities of 0%, 20% and 40% (periods 1, 5 and 2). In contrast, tax compliance is optimal for audit probabilities of 60% and 80% (periods 3 and 4). As the critical audit probability is identical for all treatments we should not expect a divergent tax evasion behavior. Above, the self-disclosure opportunity cannot be used strategically as all decision parameters are fixed and known before the tax evasion decision (last column). However, there are studies who argue that option for immunity or reduction of punishment violates the psychological contract of the actual honest

[15] For the influence of complexity and cognitive load on risk-taking see e.g., Whitney et al. (2008), Benjamin et al. (2013), Deck and Jahedi (2015), Gerhardt et al. (2016), and Fochmann and Hemmerich (2018).

[16] For the influence of legality on tax compliance see e.g., Kirchler et al. (2003), Hofmann et al. (2008) and Blaufus et al. (2016).

taxpayer and the tax authority (Feld and Frey, 2002, 2007). This leads to an erosion of tax morale and causes a crowding out of the intrinsic motivation to pay taxes. Thus, against rational predictions we should find an increased tax evasion in the Self-Disclosure Treatments in the first five periods (hypothesis 1).

For periods six to twelve, we have introduced uncertain audit probabilities: Here, we present two different audit probabilities whereby the high audit probability p_H (ranging from 40% to 80%) occurs with a probability θ of 50%. Consequently, the low audit probability p_L (ranging from 0% to 40%) occurs with a probability of $1-\theta$ of 50%. We choose the p_L and p_H in such a way, that the expected audit probability matches one of the certain audit probabilities of the first five periods. [17] We have shown above that using backward induction derives critical audit probabilities of 50% (55%) in the Self-Disclosure without (with) Penalty Treatment. If nature draws actual audit probabilities above this level, subjects should use the self-disclosure option, else not. With regard to the experiment's periods presented with uncertain audit probabilities in Table 5.1, this implies that for period 7 in which p_L amounts to 0% and p_H is 40% complete tax evasion without using the self-disclosure opportunity is always optimal independent of the actually drawn audit probability.[18] For all other periods p_L ranges between 0% and 40% and p_H amounts to either 60% or 80% so that the self-disclosure opportunity should always be used if p_H is drawn but should never be used if p_L is drawn. This is reflected in the last column.

In a second step we have calculated critical low audit probabilities of 50% (45%) in the Self-Disclosure without (with) Penalty Treatment. Up to these *low* audit probabilities complete tax compliance is optimal. As p_L fluctuates between 0% and 40%, risk-neutral and income maximizing subjects in the Self-Disclosure Treatments under uncertainty should always evade all taxes in our experimental setting and may revise their decision afterwards depending on the actually drawn audit probability. In contrast, subjects in the No Self-Disclosure Treatment should only evade all taxes if the *expected* audit probability is below 50%. In this context, we have created two periods (periods nine and twelve), in which the tax evasion strategy differs between the No Self-Disclosure Treatment and the Self-Disclosure Treatments. Here, the *low* audit probability is only 40%, thus causing a complete tax evasion strategy for the Self-Disclosure Treatments, whereas the *expected* audit probability is 60%, hence triggering tax compliance

[17] This is best reflected in Table 5.1 column 2 and 3. We have created two periods with identical audit probabilities three times (periods six and ten, eight and eleven and nine and twelve) in order to account for possible learning effects in the treatments. However, using 2-sided Mann-Whitney U tests we do not find any significant differences between pairwise period's tax evasion ratios (neither separately for each treatment nor accumulated over all treatments) that could be interpreted as learning effects.

[18] We have designed this period to stimulate high tax evasion to control for the influence of the self-disclosure opportunity where we should expect no interference.

for the No Self-Disclosure Treatment.[19] Thus, in periods with uncertain audit probabilities we expect subjects to use the self-disclosure option strategically and increase tax evasion (hypothesis 2).

As presented above, the critical audit probabilities concerning the self-disclosure usage and the tax evasion strategy differ between the Self-Disclosure Treatments under uncertainty due to the fine rate of 10% on the disclosed income in the Self-Disclosure with Penalty Treatment. However, we have set p_H and p_L so that self-disclosure usage strategies and the tax evasion strategies do not differ between the two Self-Disclosure Treatments. As theory predicts no different tax evasion behaviour we can test whether the small penalty is sufficient to decrease tax evasion in the Self-Disclosure with Penalty Treatment due to an increased complexity and the illegalization of the tax evasion (hypothesis 3).

Table 5.1: Overview of the Experiment's Periods

Period	**Audit probability**		**Expected evasion rate**		**Self-disclosure strategy**
	p_L	p_H	No Self-Disclosure Treatment	Self-Disclosure Treatments	
1	0%		100%	100%	No
2	40%		100%	100%	No
3	60%		0%	0%	No
4	80%		0%	0%	No
5	20%		100%	100%	No
6	0%	80%	100%	100%	No for p_L Yes for p_H
7	0%	40%	100%	100%	No
8	20%	60%	100%	100%	No for p_L Yes for p_H
9	40%	80%	0%	100%	No for p_L Yes for p_H
10	0%	80%	100%	100%	No for p_L Yes for p_H
11	20%	60%	100%	100%	No for p_L Yes for p_H
12	40%	80%	0%	100%	No for p_L Yes for p_H

Notes: The table presents audit probabilities per period and the respective tax evasion rates in the No Self-Disclosure Treatment (column 4) and in the Self-Disclosure Treatments (column 5). Column 6 presents the self-disclosure strategies for the two respective treatments.

[19] We create periods in which we expect to find a divergent tax compliance behavior due to the strategic use of the self-disclosure opportunity and periods where we do not expect to find any differences. This was done in order to separate the effect of the introduction of uncertain audit probabilities itself from the strategical aspect of the self-disclosure usage.

The experiment lasts about 50 minutes and participants earned 9.77 Euro on average (11.72 Euro per hour), within a range from 5.00 Euro to 13 Euro. 83 students participated in all treatments (37 females and 46 males) and are 23.7 years on average. 48.2 % study at the Faculty of Economics and Management, and 69.9% are in a Bachelor's program.[20] We find no significant differences in the individual characteristics between the treatments.

5.3.3 Variable Measurement

We measure the dependent metric variable, TAX EVASION, as ratio of the tax payments that were initially not declared to the demanded tax payments. Hence, a TAX EVASION of 1 denotes that the subject pays no taxes in the respective period, whereas a TAX EVASION of 0 expresses, that the subject pays the taxes truthfully. We additionally introduce two dichotomous variables which reflect both corner solutions that are relevant in the experiment: FULL EVADER and NON-EVADER. FULL EVADER takes the value 1 if the participant fully evades taxes in a respective period (i.e., TAX EVASION is 1) and 0 otherwise. By contrast, NON-EVADER takes the value 1 if the subject completely truthfully pays the taxes in a respective period (i.e., TAX EVASION is 0) and 0 otherwise.

Our treatments serve as independent dummy variables. SELF-DISCLOSURE WITHOUT PENALTY takes the value one if the participant is assigned to the Self-Disclosure without Penalty Treatment and zero otherwise. Likewise SELF-DISCLOSURE WITH PENALTY takes the value one if the participant is assigned to the Self-Disclosure with Penalty Treatment and zero otherwise. UNCERTAIN PERIODS is also a dummy variable and serves to distinguish periods with certain (only one probability is presented) and uncertain (two probabilities are presented) audit probabilities to reflect our within-treatment variation. Thus, UNCERTAIN PERIODS is one if the decision is made in periods 6 to 12 and is zero if the decision is made in periods 1 to 5. Additionally, we integrate the interaction term of the treatment variables and the within-treatment variation in order to control for the influence of the treatment effect in uncertain periods. Thereby, the interaction term is presented as SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS or SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS depending on the respective treatment variable in the analysis.

Besides the dummy variable UNCERTAIN PERIODS, we use another variable to test our first two hypotheses. As we only expect to find a divergent tax evasion behavior in periods 9 and 12 for risk-neutral and income maximizing subjects where the low audit probability amounts to 40% and the high audit probability amounts to 80%, we generate the dichotomous variable EXPECTED DIFFERENCE which takes the value 1 for periods 9 and 12 and zero for all other periods. Hence, even if we do not find a general effect of uncertain audit probabilities on tax evasion behavior, we expect the effect to be noticeable for these two periods. Consequently, we also use the respective interaction terms to analyze

[20] Please note that prior literature has shown that student decisions serve as a good surrogate for non-student decisions and do not significantly differ from each other. See for example Ashton and Kramer (1980), Remus (1996), Elliot et al. (2007), Liyanarachchi (2007), Depositario et al. (2009) and Alm et al. (2015).

the influence of the self-disclosure opportunities for these two periods in which we expect to find a different tax evasion behavior.

Additionally, we want to test, whether subjects adapt their tax evasion behavior in the course of time (e.g., through learning or experience). However, as the periods are highly correlated with the dummy variable UNCERTAIN PERIODS by design, we adjust this variable so that we measure adaptions through learning for both risk segments separately. We do that by starting to count the periods again when the uncertain audit probabilities are introduced. Thus, for each subject we find twice the round 1 (in period 1 and period 6), twice the round 2 (in period 2 and period 7) and so on. Hence, we are able to measure adaption mechanism while taking the introduction of uncertainty in the course of the experiment into account and call the variable for reasons of simplifications ROUND. AUDIT PROBABILITY is also an experiment specific measure and presents the respective audit probability under which the tax evasion decision is made. For periods with two possible audit probabilities, AUDIT PROBABILITY reflects the expected value of both probabilities.

As further control variables, we use subject specific characteristics, such as AGE (measured in years), gender (FEMALE which takes the value 1 if the subject is female and 0 otherwise) and their field of study (ECONOMICS AND MANAGEMENT which takes the value 1 if the participants studies at this faculty and zero otherwise). We measure RISK AVERSION based on the respective SOEP question by the self-assessment of the participants who are asked to report their risk attitude on an 11-point scale with 0 = "I am not risk-loving at all" and 10 = "I am very risk-loving". TAX EXPERIENCE presents the subject's experience with filing tax returns and takes the value 0 if the subject either reports to have not yet filed such a return or cannot remember whether she already did so. It takes the value 1 (2, 3) if the subject already once (twice, more than twice) filed a tax return. TAX MORALE presents the subject's self-reported tax morale on a 10-point scale when being asked whether it is okay to evade taxes if you have the opportunity to do so. While 0 mirrors subjects with low self-reported tax morale ("It is all right to evade taxes in any cases"), 9 reflects the highest tax morale ("You are not allowed to evade taxes on any terms"). POSTIVE (NEGATIVE) RECIPROCITY presents the subject's answer to the question: "If somebody does something good to you, are you content to reciprocate even if it was not agreed upon before? (If someone treats you badly, do you also treat this person badly?)". On an 11-point scale subjects rate whether they agree on no account (0 points) or any account (10 points) with this statement. Hence, the higher the statement, the more subjects are affected by the actions of others and adapt their counteractions.[21] COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems). On an 11-

[21] We measure reciprocity in order to control for the subject's affection towards the psychological contract between taxpayer and state. If the subjects indicates that reciprocity plays a role for her, a violation of the psychological contract due to the self-disclosure opportunity may decrease tax morale which can cause a crowding out of intrinsic motivated taxpayers, see e.g., Feld and Frey (2002, 2007).

point scale subjects could rate their opinion, whereby 0 presents "not complex at all" ("very unfair") and 10 "very complex" ("very fair").[22]

5.4 Results

5.4.1 Descriptive Statistics

We start examining tax compliance behavior for all three treatments by analyzing TAX EVASION. The tax evasion rates per period are presented in Table 5.2. While columns 1 and 2 display the periods and respective audit probabilities, column 3 displays the respective expected tax evasion rate under the given experiment's parameters assuming risk-neutral and income maximizing subjects. As presented, subjects should either evade no or full taxes independent of the treatment. The only exceptions are periods 9 and 12, in which subjects in the No Self-Disclosure Treatment should not evade any taxes at all, whereas subjects in the Self-Disclosure Treatments should fully evade taxes. Column 4 (5; 6) reveals the tax evasion results for the No Self-Disclosure Treatment (Self-Disclosure without Penalty Treatment; Self-Disclosure with Penalty Treatment).

Tax evasion over all periods amounts to 38.44% if there is no option on self-disclosure and is 47.91% (40.88%) if the option on self-disclosure comprises no (a) penalty. Thereby the total tax evasion rate in the Self-Disclosure without Penalty Treatment is significantly higher than in both other treatments ($p = 0.0183$ and $p = 0.0532$ respectively, 2-sided Mann-Whitney U tests, see columns 7 and 9). Therefore, we continue analyzing the experiment with regard to the factors that might cause the increase of tax evasion. First, we analyze periods with certain audit probabilities to control whether we find an erosion of tax morale that causes the decrease of tax compliance. Second, we analyze periods with uncertain audit probabilities to test whether subjects strategically use the self-disclosure opportunity which increases tax evasion.

As described before, the experiment starts with certain audit probabilities in the first five periods. Given that all other parameters are fixed too, the erosion of tax morale caused by the violation of the psychological contract between taxpayer and tax authority due to the introduction of the self-disclosure opportunity is the only factor than can decrease tax morale (hypothesis 1). However, our results reveal that tax evasion in the first five periods (with certain audit probabilities) nearly amounts to the same level with 48.30% for the No Self-Disclosure Treatment, 47.11% for the Self-Disclosure without Penalty Treatment and 47.12% in the Self-Disclosure with Penalty Treatment. These differences are not significant so that we may not confirm our first hypothesis.

For periods with uncertain audit probabilities, we expect subjects to tactically integrate their self-disclosure option into their tax evasion decision under uncertain audit probabilities which should lead

[22] The question on the fairness of the contribution's and control's system also aims at testing the participant's evaluation of the tax system and thus the psychological contract.

to increased tax evasion if such an option is offered (hypothesis 2). We have explained that given our audit probabilities' parameters in the experiment and the assumption of risk-neutral and income maximizing subjects, tax evasion strategies should only differ in periods 9 and 12. While risk-neutral subjects in the No Self-Disclosure Treatment should not evade any taxes, subjects in the Self-Disclosure Treatments should evade all taxes (see Table 5.2, column 3). In period 9, the tax evasion level of subjects in the No Self-Disclosure Treatment (2.82%) is significantly lower than the tax evasion level in the Self-Disclosure without Penalty Treatment (21.04%, $p = 0.0100$) and in the Self-Disclosure with Penalty Treatment (14.40%, $p = 0.0429$). In period 12, which displays the same audit probability lottery as period 9, tax evasion has increased to 8.83% in the No Self-Disclosure Treatment and to 17.16% in the Self-Disclosure with Penalty Treatment. This difference is not significant ($p = 0.2194$). However, tax evasion in the Self-Disclosure without Penalty Treatment amounts to 34.58% which is significantly higher than in both other treatments ($p = 0.0046$ and $p = 0.0845$ respectively). Analyzing the tax evasion rates of both periods accumulated, we find that the tax evasion rate in the No Self-Disclosure Treatment amounts to 5.83%, whereas it is 27.81% (15.78%) in the Self-Disclosure without (with) Penalty Treatment. These differences between the No Self-Disclosure Treatment and both Self-Disclosure Treatments are highly significant with p-values of 0.0001 and 0.0232 respectively. Thus, for periods in which we explicitly expect to find a different tax evasion behavior, we indeed find that the voluntary disclosure opportunity is strategically integrated into the decision process and increases the tax evasion level given initially uncertain audit probabilities. Hence, we may confirm our second hypothesis.

Additionally, we also test whether tax evasion rates increase in other periods with uncertain audit probabilities than in the expected two. While we find that tax evasion in the Self-Disclosure without Penalty Treatment is also significantly higher than in the No Self-Disclosure Treatment in periods 10 and 11 ($p = 0.0085$ and $p = 0.0928$ respectively), we do not find that the introduction of a self-disclosure opportunity with penalty payments increases tax evasion rates significantly in any other period. In line with these findings, we analyze the overall tax evasion rates in periods with uncertain audit probabilities. In periods 6 to 12, tax evasion amounts to 31.40% in the No Self-Disclosure Treatment and to 48.48% (36.43%) in the Self-Disclosure without (with) Penalty Treatment. Granting impunity for the self-disclosure option significantly increases tax evasion in comparison to both, the No Self-Disclosure Treatment ($p = 0.0004$) and the Self-Disclosure with Penalty Treatment ($p = 0.0064$). By contrast, although tax evasion under a self-disclosure with penalty option is 16% higher than without the option, this difference is not statistically significant ($p = 0.3922$). Thus, linking a self-disclosure opportunity with a reduced penalty payment is sufficient to limit the increase of tax evasion which confirms our third hypothesis.

For the sake of completeness we also test whether we find a significant different tax evasion behavior if we cumulate all other periods with the same expected tax evasion rate (i.e., either 0% or 100%) with regard to certain and uncertain periods. We do not find any significant differences for periods with

certain probabilities: For periods with an expected tax evasion rate of 0% (periods 3 and 4) the p-value amounts to 0.2278 (0.7312) when comparing the No Self-Disclosure Treatment with the Self-Disclosure without (with) Penalty Treatment. For periods with an expected tax evasion rate of 100% (periods 1, 2 and 5) the respective p-values amount to 0.2153 and 0.5627. When testing for periods with uncertain audit probabilities and an expected tax evasion rate of 100% for both treatments (periods 6, 7, 8, 10, and 11) we do not find significant differences between the No Self-Disclosure Treatment and the Self-Disclosure with Penalty Treatment ($p = 0.8030$). However, tax evasion rates in the Self-Disclosure without Penalty Treatment are significantly higher than in the No Self-Disclosure Treatment ($p = 0.0080$) and in the Self-Disclosure with Penalty Treatment ($p = 0.0136$). These results underpin our findings described when analyzing periods with uncertain audit probabilities. Thus, we may confirm our second and third hypothesis.

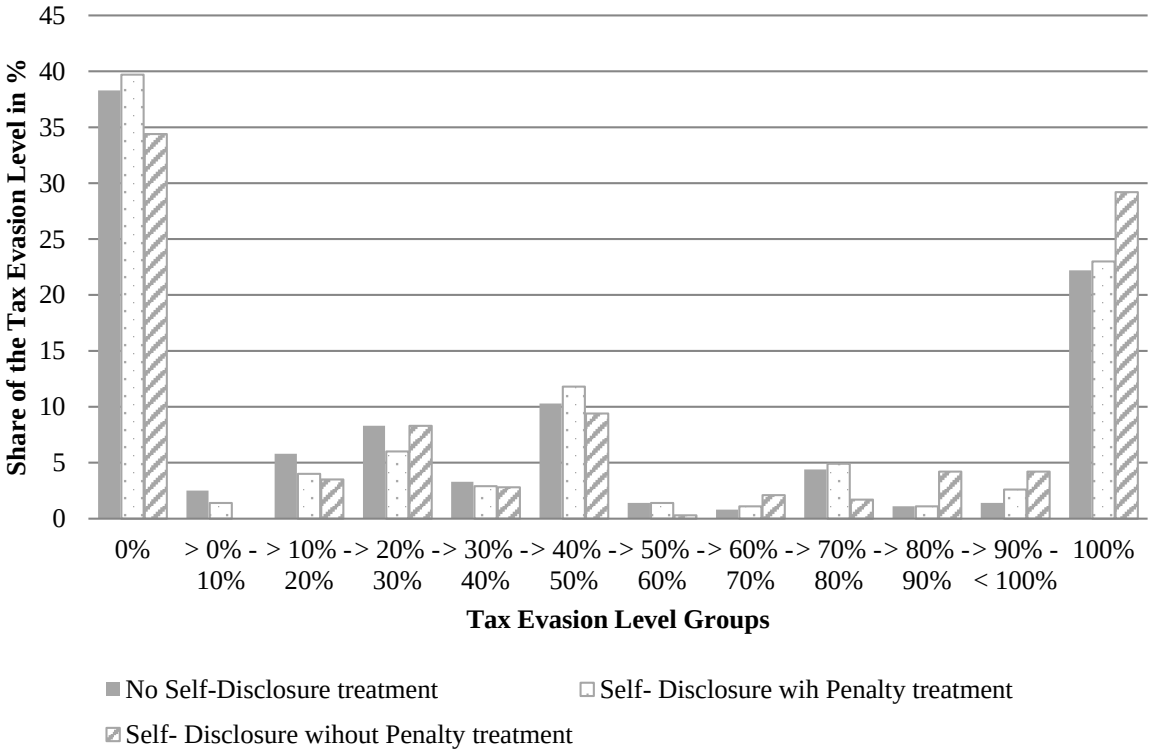
Table 5.2: Tax Evasion Rates per Period

Period	Audit probability	Expected evasion rate	No Self-Disclosure (1)	Self-Disclosure without Penalty (2)	Self-Disclosure with Penalty (3)	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Total			38.44%	47.91%	40.88%	0.0183	0.6703	0.0532
1-5			48.30%	47.11%	47.12%	0.7267	0.8082	0.8641
6-12			31.40%	48.48%	36.43%	0.0004	0.3922	0.0064
9&12	40% / 80%	0%/100%*	5.83%	27.81%	15.78%	0.0001	0.0232	0.0797
1	0%	100%	92.42%	86.20%	96.98%	0.2920	0.7054	0.1402
2	40%	100%	51.63%	46.51%	45.34%	0.6466	0.5542	0.9418
3	60%	0%	15.04%	26.41%	20.22%	0.4372	0.8779	0.6031
4	80%	0%	5.33%	13.85%	6.07%	0.2422	0.4865	0.5748
5	20%	100%	77.08%	62.60%	66.97%	0.2046	0.2212	0.7881
6	0% / 80%	100%	40.08%	54.98%	38.71%	0.2204	0.9628	0.1458
7	0% / 40%	100%	76.79%	63.96%	66.90%	0.2185	0.2810	0.8750
8	20% / 60%	100%	25.08%	43.39%	35.42%	0.1736	0.6479	0.4140
9	40% / 80%	0%/100%*	2.82%	21.04%	14.40%	0.0100	0.0429	0.4716
10	0% / 80%	100%	34.53%	68.54%	44.56%	0.0085	0.5776	0.0443
11	20% / 60%	100%	31.64%	52.92%	37.84%	0.0928	0.6296	0.1819
12	40% / 80%	0%/100%*	8.83%	34.58%	17.16%	0.0046	0.2194	0.0845

Notes: This table displays tax evasion rates per period and treatment. The first column presents the period(s), the second column the respective audit probabilities. The third column presents the expected tax evasion rate assuming risk-neutrality. In periods 9 and 12 the * denotes that for the Self-Disclosure Treatments the optimal tax evasion rate is 100%, whereas is 0% for the No-Self Disclosure Treatment. The fourth (fifth; sixth) column displays the respective tax evasion rates in the No Self-Disclosure Treatment (Self-Disclosure without Penalty Treatment; Self-Disclosure with Penalty Treatment). Using 2-sided pairwise Mann-Whitney U tests we analyze whether the tax evasion rates significantly differ between the treatments. The respective p-values are presented in column 7 (8;9) for the comparison between the No Self-Disclosure Treatment and the Self-Disclosure without Penalty Treatment (No Self-Disclosure Treatment and Self-Disclosure with Penalty Treatment; Self-Disclosure without Penalty Treatment and Self-Disclosure with Penalty Treatment).

Figure 5.2 illustrates the distribution of different tax evasion level groups for all treatments. The first three bars present the share of non-evading decisions over all periods. The last three bars reflect the fraction of the Full Evader over all periods. Every other bar displays a group of different tax evasion levels in a 10%-range, starting with the share of decisions in which subjects evade more than 0% up to 10%. The figure reveals that the distribution of the tax evasion levels is nearly identical, especially for the No Self-Disclosure Treatment and the Self-Disclosure with Penalty Treatment. The noticeable exception is the share of Non-Evader and Full Evader in the Self-Disclosure without Penalty Treatment compared to both other treatments.

Figure 5.2: Distribution of Tax Evasion Levels by Treatment



Note: This figure illustrates the distribution of different tax evasion level groups for all treatments. The first three bars present the share of non-evading decisions over all periods. The last three bars reflect the fraction of the Full Evader over all periods. Every other bar displays a group of different tax evasion levels in a 10%-range, starting with the share of decisions in which subjects evade more than 0% up to 10%.

We analyze the occurrence of NON-EVADER and FULL EVADER in all treatments by running the 2-sided Fisher's exact test.[23] Thereby, the analysis of Non-Evaders underpins our results obtained when analyzing the tax evasion rates. We do not find an erosion of tax morale caused by the self-disclosure opportunity (rejection of hypothesis 1), but we find that subjects use the self-disclosure strategically in periods with uncertain audit probabilities by increasing their tax evasion (confirmation of hypothesis 2). Again, this increase is limited to the two expected periods if the self-disclosure opportunity is linked with a penalty payment (confirmation of hypothesis 3). Analyzing Full Evaders, we also do not find a

[23] The tests are presented in Table 5.5 and Table 5.6 in Section 5.11 (Appendix E).

significant difference in the share of full evading decisions between the three treatments. Thus we again do not find a decrease of tax morale and a crowding out of formally intrinsic subjects and must reject hypothesis 1. Analyzing the strategic integration of the self-disclosure option into the tax evasion decision, we do not find significant differences in the two periods where we expect them, but find that the share of Full Evaders is significantly higher in the Self-Disclosure without Penalty Treatment compared to both other treatments. Thus, we may also confirm hypotheses 2 and 3.

5.4.2 Multivariate Analysis

We run several panel regressions to test our results obtained above and control whether other independent variables influence the tax evasion behavior. We run random-effects panel regressions using TAX EVASION as dependent variable to exploit the panel structure of our data and cluster on subject's level to account for the dependence of the twelve decisions made by one subject.[24] We always run two different panel regressions which on the one hand test the pure influence of the introduction of the self-disclosure opportunities as well as of uncertain audit probabilities and on the other hand take further control variables into consideration. As already presented in the bivariate analyses, we test our hypotheses twofold: First, we analyze whether the self-disclosure options lead to an increased tax evasion behavior in those two periods (9 and 12) in which we indeed expect participants to show a divergent tax evasion behavior. Second, we explore whether the introduction of uncertainty itself also influenced tax evasion decisions.

In a first step, we test for treatment effects by integrating the dummy variables SELF-DISCLOSURE WITHOUT PENALTY and SELF-DISCLOSURE WITH PENALTY into the regression equation. The dummies take the value 1 if the subject is in the respective Self-Disclosure Treatment and 0 otherwise. The No Self-Disclosure Treatment serves as control group. Furthermore, we use the dummy variable EXPECTED DIFFERENCE which takes the value 1 for periods 9 and 12 and zero otherwise. The additional interaction terms with the respective treatments SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE are used to analyze the influence of the Self-Disclosure Treatments on the tax evasion in periods 9 and 12. Equation 5.9 presents the respective random-effects panel regression:

[24] As illustrated in Figure 5.2, we find two corner solution responses, i.e. for tax evasion amounting to zero and one. To take account of this, we have run all regressions with TAX EVASION also by running tobit panel regressions as well as tobit regressions while cluster on subject's level. The results are robust in all regressions.

$$\begin{aligned}\text{TAX EVASION}_{it} = {} & \alpha + \beta_1 \text{SELF} - \text{DISCLOSURE WITHOUT PENALTY}_i \\ & + \beta_2 \text{SELF} - \text{DISCLOSURE WITH PENALTY}_i + \beta_3 \text{EXPECTED DIFFERENCE}_t \\ & + \beta_4 \text{SELF} - \text{DISCLOSURE WITHOUT PENALTY}_i \text{ X EXPECTED DIFFERENCE}_t \\ & + \beta_5 \text{SELF} - \text{DISCLOSURE WITH PENALTY}_i \text{ X EXPECTED DIFFERENCE}_t \\ & + \sum_{k=1}^{l} \beta_k Controls_i + \varepsilon_{it} + u_i\end{aligned} \tag{5.9}$$

where $i = 1, \dots, 83$ and $t = 1, \dots, 12$. We use the control variables which are already presented in Section 5.3.3. In a second step, we test whether the introduction of the self-disclosure options increases tax evasion even more than expected, i.e., in periods with uncertain audit probabilities in general. Thus, besides our treatment variable we introduce UNCERTAIN PERIODS which takes the value 1 if the tax evasion decision is made under uncertain audit probabilities (i.e., in the periods six to twelve). The additional interaction terms with the respective treatments SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS are used to analyze the general influence of the Self-Disclosure Treatments on tax evasion in uncertain periods. Equation 5.10 presents this random-effects panel regression:

$$\begin{aligned}\text{TAX EVASION}_{it} = {} & \alpha + \beta_1 \text{SELF} - \text{DISCLOSURE WITHOUT PENALTY}_i \\ & + \beta_2 \text{SELF} - \text{DISCLOSURE WITH PENALTY}_i + \beta_3 \text{UNCERTAIN PERIODS}_t \\ & + \beta_4 \text{SELF} - \text{DISCLOSURE WITHOUT PENALTY}_i \text{ X UNCERTAIN PERIODS}_t \\ & + \beta_5 \text{SELF} - \text{DISCLOSURE WITH PENALTY}_i \text{ X UNCERTAIN PERIODS}_t \\ & + \sum_{k=1}^{l} \beta_k Controls_i + \varepsilon_{it} + u_i\end{aligned} \tag{5.10}$$

Table 5.3 presents the random-effects panel regressions' results with TAX EVASION as the dependent variable. We present the results of the treatment effects whereby the No Self-Disclosure Treatment serves as reference group. In the first two models, we use the treatments and the dummy variable EXPECTED DIFFERENCE and the respective interaction terms to analyze whether the introduction of the Self-Disclosure Treatments leads to a divergent tax evasion behavior in the two periods where we expect such a different behavior. In models 3 and 4, we use the treatment variables, the dummy variable UNCERTAIN PERIODS as well as the interaction terms of these dummy variables to study whether the existence of a self-disclosure opportunity influences tax compliance facing certain and/or uncertain audit probabilities. Additionally, we add ROUND to control for behavioral adaptions over time and we control for the influence of AUDIT PROBABILITY.

Table 5.3: Regression Results for Tax Evasion

TAX EVASION	(1)	(2)	(3)	(4)
SELF-DISCLOSURE	0.0697	0.0595	-0.0119	-0.0221
WITHOUT PENALTY	(0.0634)	(0.0746)	(0.0627)	(0.0735)
SELF-DISCLOSURE	0.0094	0.0213	-0.0118	0.0001
WITH PENALTY	(0.0542)	(0.0475)	(0.0493)	(0.0442)
EXPECTED DIFFERENCE	-0.1511***	-0.1511***		
	(0.0392)	(0.0394)		
SELF-DISCLOSURE WITHOUT	0.1502**	0.1502**		
PENALTY X EXPECTED DIFFERENCE	(0.0619)	(0.0622)		
SELF-DISCLOSURE WITH	0.0901*	0.0901*		
PENALTY X EXPECTED DIFFERENCE	(0.0501)	(0.0503)		
UNCERTAIN PERIODS			-0.1365***	-0.1365***
			(0.0319)	(0.0320)
SELF-DISCLOSURE WITHOUT			0.1827***	0.1827***
PENALTY X UNCERTAIN PERIODS			(0.0522)	(0.0525)
SELF-DISCLOSURE WITH			0.0621	0.0621
PENALTY X UNCERTAIN PERIODS			(0.0428)	(0.0430)
ROUND	0.0010	0.0010	0.0002	0.0002
	(0.0055)	(0.0055)	(0.0050)	(0.0050)
AUDIT PROBABILITY	-1.1027***	-1.1027***	-1.1438***	-1.1438***
	(0.0507)	(0.0509)	(0.0520)	(0.0522)
AGE		-0.0058		-0.0058
		(0.0066)		(0.0066)
FEMALE		-0.0638		-0.0638
		(0.0544)		(0.0544)
ECONOMICS AND MANAGEMENT		0.0075		0.0075
		(0.0519)		(0.0519)
RISK AVERSION		-0.0235**		-0.0235**
		(0.0109)		(0.0109)
TAX EXPERIENCE		0.0222		0.0222
		(0.0220)		(0.0220)
TAX MORALE		0.0026		0.0026
		(0.0099)		(0.0099)
POSITIVE RECIPROCITY		0.0082		0.0082
		(0.0116)		(0.0116)
NEGATIVE RECIPROCITY		0.0046		0.0046
		(0.0099)		(0.0099)
FAIRNESS		-0.0043		-0.0043
		(0.0086)		(0.0086)
COMPLEX DECISION		0.0179*		0.0179*
		(0.0093)		(0.0093)
CONSTANT	0.8654***	1.0185***	0.9400***	1.0931***
	(0.0397)	(0.1944)	(0.0427)	(0.1952)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000

Notes: The table presents the results of random-effects panel regressions which are clustered on subject level. The metric variable TAX EVASION serves as dependent variable and presents the ratio of tax payments that were initially not declared to the demanded tax payments. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. ROUND (AUDIT PROBABILITY) presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale

and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Models 1 and 2 reveal that we do not find any treatment effect on the level of tax evasion. Thus, we may not confirm our first hypothesis that the self-disclosure opportunity leads to morale erosion which increases tax evasion. In periods 9 and 12 tax evasion significantly decreases by 15 percentage points in comparison to all other periods which is not surprising as these two periods present the highest expected audit probabilities under uncertainty. However, the analysis of the interaction terms is far more interesting to test our second hypothesis. In accordance with rational predictions assuming risk-neutral subjects, we find that the Self-Disclosure without (with) Penalty opportunity increases tax evasion (slightly) significantly in periods 9 and 12. In periods 9 and 12, tax evasion is 15 (9) percentage points higher in the Self-Disclosure without (with) Penalty Treatment than in the No Self-Disclosure Treatment. Thus, we may confirm hypothesis 2. We also run Wald tests after each regression. Hereby, we test whether the two coefficients of the two Self-Disclosure Treatments and of the two respective interaction terms differ significantly. For models 1 and 2, we neither find any significant treatment effects ($p = 0.3925$ and $p = 0.6068$, respectively) nor an interaction effect ($p = 0.3114$ and $p = 0.3139$, respectively).

In contrast to the first two models, models 3 and 4 analyze the impact of the self-disclosure opportunity on tax evasion for all periods with uncertain audit probabilities, but not only for periods 9 and 12. Here, we also find that none of the two self-disclosure opportunities has an effect on the level of tax evasion, so that we again may not confirm the first hypothesis. Whereas the interaction term SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS is positive and highly significant, the interaction term of the Self-Disclosure with Penalty Treatment is not. Therefore, we may conclude that offering the self-disclosure option with impunity leads to an excessive increase of tax evasion which is not limited to rational prediction assuming risk-neutral subjects, whereas the self-disclosure option with penalty payments limits the tax evasion' s increase. These findings are also supported by the Wald tests. While we again find no significant differences between both self-disclosure Treatments over all periods, we find a significant difference in the interaction terms ($p = 0.0183$ for model 3 and $p = 0.0189$ for model 4). Thus, tax evasion under uncertain audit probabilities is significantly higher in the Self-Disclosure without Penalty Treatment than in the Self-Disclosure with Penalty Treatment. This confirms our third hypothesis that small penalty payments which accompany the self-disclosure option are sufficient to limit tax evasion.

Over all models, ROUND does not significantly influence tax evasion so that we do not find an adaption behavior in the course of time. AUDIT PROBABILITY however influences tax compliance significantly. In models 2 and 4 we integrate subject specific control variables and find that RISK AVERSION has a negative significant effect on tax evasion. The riskier subjects describe themselves, the higher is their tax evasion level. Additionally, COMPLEX DECISION also significantly affects tax

evasion. Subject who state that it was a complex decision whether and how much to evade, tend to evade more taxes.

Besides the random-effects panel regressions, we run random-effects logit regressions using the dichotomous variables NON-EVADER and FULL EVADER as dependent variable. For both regressions, we also cluster on the subject's level to account for the subject's dependence within the twelve decision periods. Again, we start by analyzing the effect of periods 9 and 12 where we expect to find the divergent tax evasion behavior and continue to test whether the introduction of uncertainty itself leads to a different tax evasion behavior. We run these regressions twice, first analyzing the pure treatment and uncertain periods' factors and second by integrating further control variables. The results for NON-EVADER are displayed in Table 5.7 and reveal that all results obtained above may be approved. We may not confirm hypothesis 1 that the introduction of the self-disclosure opportunity leads to an erosion of tax morale, thus reducing tax compliance. However, we may confirm hypothesis 2 and 3 that the self-disclosure option is used strategically thus increasing tax evasion under uncertain audit probabilities. However, this increase can be limited by relating the option to a small fine. The results for FULL EVADER are presented in Table 5.8. As for all other tests, we may not confirm our first hypothesis. In contrast to the other regressions, we may only confirm our second hypothesis for the analysis including all periods with uncertain audit probabilities (models A7 and A8), but not for those two periods where we actually expect a tactical use of the self-disclosure opportunity to result in different tax evasion strategies. Again, we may also confirm hypothesis three in those two models.

5.4.3 Robustness Analysis

We run several robustness checks in order to test whether the results obtained above are stable. First, we rerun all regressions without ROUND in order to control for probable distortions due to the repeated counting of rounds in periods with certain and uncertain audit probabilities. As we do not find any learning effects in the course of the experiment, omitting ROUND is uncritical. Table 5.9 to Table 5.11 in Section 5.11 (Appendix E) present the respect regression results. Comparing these with the regressions analyzed above, we do not find any considerable differences. Furthermore, we also integrate LAST PERIOD AUDIT which is a dummy variable and takes the value 1 if the subject was audited in the previous period and 0 otherwise. By doing so, we want to test whether we also find a "bomb crater effect".[25] The bomb crater effect describes subjects to increase tax evasion after they were audited in the previous period. The results are presented in Table 5.12 in Section 5.11 (Appendix E) and reveal that we do not find a bomb crater effect.[26] Our results obtained above remain unchanged. As we lose one

[25] Early paper on the bomb crater effect are e.g., Mittone (2006), Maciejovsky et al. (2007) and Kastlunger et al. (2009).

[26] Unfortunately, we are only able to rerun the regressions for our metric tax evasion variable. For both dummy variables, NON-EVADER and FULL EVADER, we were not able to receive results when adding the variable LAST PERIOD AUDIT and while clustering at subject's level. We could have presented results for adding the variable without clustering at subject's level. However, as these results are not automatically comparable to our initial regressions, we refrain from analyzing and presenting them.

observation per subject because the first period is skipped as LAST PERIOD AUDIT may not take any value there, we decide to run further tests without this variable.

For both robustness analyses, we may not confirm hypotheses 1 that the introduction of the self-disclosure opportunity leads to a crowding out of formerly honest taxpayers due to the erosion of tax morale. However, we may confirm the second hypothesis that the self-disclosure opportunity is used strategically under uncertain audit probability which leads to an increase in tax evasion. However, this increase may be limited effectively by combining the self-disclosure opportunity with a penalty payment (hypothesis 3).

5.5 Additional Analyses

We want to profoundly investigate our results obtained above related to our three hypotheses. First, following our first hypothesis we want to investigate whether we are able to find an erosion of tax morale by analyzing in the first periods more closely. Second, we investigate whether the self-disclosure option is indeed used strategically to acknowledge our second hypothesis. Third, we analyze why we find an increased tax evasion behavior in the Self-Disclosure without Penalty Treatment in comparison to the Self-Disclosure with Penalty Treatment although rational choice predicts no differences (third hypothesis).

5.5.1 Tax Morale

As presented in Section 5.4, we do not find evidence of a crowding out of formerly intrinsic motivated subjects due to the introduction of a self-disclosure opportunity (hypothesis 1). Therefore, we want to analyze subjects' tax morale more closely. First, we consider the subject's self-reported TAX MORALE that subjects indicate in the questionnaire on a 10-point scale when being asked whether it is okay to evade taxes if you have the opportunity to do so.[27] Subjects in the No Self-Disclosure Treatment rate their tax morale with 6.93 on average whereas subjects in the Self-Disclosure without (with) Penalty Treatment rate it with 6.29 (7.31). We do not find any significant differences between any treatment comparisons.

However, TAX MORALE may be distorted due to a wrong self-perception (e.g., Bem, 1972; John and Robins, 1994), the knowledge of social desirability (e.g., Randall and Fernandes, 1991; Andreoni and Erard, 1998; Wenzel, 2005; Bobek et al., 2013) or the experiment's influence on the post-experimental questionnaire (Weimann and Brosig-Koch, 2019). Therefore, we analyze the tax evasion decision in the first period in order to get an impression of the subject's actual tax morale. We have implemented the first period with an audit probability of zero percent in order to test for the intrinsic motivation to pay taxes. Independent of their own risk attitude, rational subjects should evade all taxes in this period as it

[27] In contrast to the questionnaire and in line with all previous tests, TAX MORALE is analyzed inversely so that a higher value presents a higher tax morale.

may not be detected by design. Subjects who do not evade (all) taxes have to be intrinsically motivated to do so, best explained with a high tax morale.[28] Analyzing the first period, we find that subjects evade 92.42% in the No Self-Disclosure Treatment and 86.20% (96.98%) in the Self-Disclosure without (with) Penalty Treatment. Although hypothesis 1 predicts a crowding out of tax morale due to the self-disclosure option, we do not find significant differences between the treatments.[29] Analyzing the first period more precisely, in the No Self-Disclosure Treatment (Self-Disclosure without Penalty Treatment; Self-Disclosure with Penalty Treatment) 6.67% (8.33%; 0%) evade no taxes, 0% (4.2%; 0%) evade 25%, 0% (4.2%; 0%) evade 50%, 0% (0%; 3.45%) evade 62.5%, 3.33% (0%; 6.90%) evade 75%, and 3.45% (8.33%; 0%) evade more than 90% of the taxes. Hence, 86.67% (75%; 89.66%) evade all taxes.[30] These results reveal that we do not find any significant differences regarding intrinsic motivation between the treatments and that we do not find many subjects who actually face an intrinsic motivation to pay taxes. Thus, we again find no evidence for the self-disclosure option to cause an erosion of tax morale.

5.5.2 The Use of the Self-Disclosure Option

Besides the tax evasion analysis, we investigate the use of the self-disclosure opportunity in order to evaluate whether the usage of the self-disclosure option increases significantly when periods with uncertain audit probabilities are introduced. Due to unchanged environmental parameters we expect subjects in periods with certain audit probabilities not to use the Self-Disclosure option unless they underlie an altered risk-perception. In contrast, we expect subjects to use this opportunity strategically in uncertain periods if the high audit probability is drawn.[31]

Thereby, we only analyze the self-disclosure choices which follow a tax evasion decision above zero, i.e., the subject has actually evaded taxes and the realistic opportunity to revise her decision is given. We find that subjects revise their decision in 9.33% (21.74%) of all possible cases under certain audit probabilities and in 27.19% (31.36%) under uncertainty in the Self-Disclosure without (with) Penalty Treatment. While the difference in the Self-Disclosure without Penalty Treatment is highly significant ($p = 0.0026$, 2-sided Fisher's exact test), it is not significant in the Self-Disclosure with Penalty Treatment ($p = 0.159$). Especially the frequent use of the self-disclosure option in the first five periods in the Self-Disclosure with Penalty Treatment might indicate that subjects feel guilty after evading taxes thus revising their decision (see Section 5.5.3). Besides, we also analyze the use of the self-disclosure option for the Full Evader exclusively. For Full Evader we find that 5.71% (6.67%) of all choices are revised under certain audit probabilities while it is 28.57% (25.71%) of all full evading choices under

[28] We implemented the audit probability of zero percent in the first period to prevent distortions by previous periods' results and experiences.

[29] For a presentation of the respective p-values, see Table 5.2.

[30] Again, these differences are not significant, see Table 5.6.

[31] Period 7 is an exception with relatively low audit probabilities of 0% and 40%. Here, we do not expect subjects to use the tax evasion decision at all independent of the actual drawn audit probability.

uncertain audit probabilities in the Self-Disclosure without (with) Penalty Treatment. This difference is significant with a p-value of 0.010 (0.026).

As our analyses reveal different results we also run random-effects logit regressions using SELF-DISCLOSURE USE as dependent variable. SELF-DISCLOSURE WITH PENALTY, UNCERTAIN PERIODS, SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS, ROUND and AUDIT PROBABILITY serve as independent variables with which we test whether the use of the self-disclosure opportunity rises under uncertainty and differently between the two treatments while taking the respective round and audit probability into consideration. In a second step, we also take the presented subject specific variables into consideration. Again, we only run the regression for those decisions in which the subject initially decided to evade taxes (i.e., decisions with a tax evasion rate above 0%) and present the results in Table 5.13. While models A25 and A26 take all relevant observations into account, models A27 and A28 only test for full evading decisions. The results reveal that the use of the self-disclosure option increases significantly when uncertain audit probabilities are introduced. Additionally, the self-disclosure use is higher in the Self-Disclosure with Penalty Treatment due to the high number of uses in the first five periods. This is also underpinned by the negative interaction term. Additionally, the probability to actually use the offered self-disclosure option decreases the more rounds pass by and increases the higher the actual audit probability is. When analyzing full evading decisions we only find a significant influence of uncertain periods on the self-disclosure use. These results also confirm the bivariate analyses that we do not find a different self-disclosure use between the treatments but when uncertain audit probabilities are introduced.

As we have demonstrated in Section 5.3.1 we expect participants in the Self-Disclosure Treatments to fully evade taxes in periods with uncertain audit probabilities and to self-disclose their evaded taxes if the high audit probability is drawn. Period 7 is an exception as the high audit probability is only 40% so that we expect subjects not to use the self-disclosure option in this period. Hence, we analyze the self-disclosure use in dependence of the actual drawn audit probability for subjects who have actually evaded taxes excluding period 7. We find that 9.4% of the subjects use the option if the low audit probability is drawn, while 50.5% use the option when the high audit probability is drawn. This difference is highly significant with a p-value of 0.0000 (2-sided Fisher's exact test).[32] This result is confirmed by the evaluation of Full Evader exclusively: While 10% revise their full evading decisions if the low audit probability is drawn, 51.5% revise it if the high audit probability is drawn. This difference is also significant ($p = 0.0000$).

Again, we run random-effects logit regressions to test our findings and use SELF-DISCLOSURE USE as dependent variable. We now integrate our new variable HIGH AUDIT PROBABILITY which takes the value 1 if the high audit probability is drawn and 0 if the low audit probability is drawn. However,

[32] The results remain unchanged if we test for both Self-Disclosure treatments separately.

as this variable is only relevant for periods with uncertain audit probabilities we have to drop UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS for these analyses. The results are presented in Table 5.14, again first shown for all relevant observations in models A29 and A30 as well as for full evading decisions exclusively in models A31 and A32.[33] The analyses reveal that the self-disclosure option is used strategically when the high audit probability is drawn. Thereby, we do not find any treatment effects.

Summarizing, the results indicate that subjects, especially in the Self-Disclosure without Penalty Treatment, use the self-disclosure option rational, i.e., when they face uncertain audit probabilities. Additionally, participants in both treatments use the option strategically, i.e., after the high audit probability is drawn in uncertain periods.

5.5.3 Complexity and Illegalization

As we have demonstrated in section 5.3.1 we have modelled the two Self-Disclosure Treatments in such a way that they only differ in the fine's level which is levied on the evaded taxes if participants use the self-disclosure opportunity. While subjects in the Self-Disclosure without Penalty Treatment face impunity, subjects in the Self-Disclosure with Penalty Treatment pay a fine of 10% of the disclosed taxes. Thus, while risk-neutral subjects in the Self-Disclosure without Penalty Treatment are expected to evade all taxes as long as the audit probability is below 50%, subjects in the Self-Disclosure with Penalty Treatment should fully evade up to an audit probability of 45%. However, as presented in Table 5.2 the audit probabilities in each period are set in 20% steps so that subjects may face an audit probability of 40% and 60%, but not between 45% and 50%. Thus, we may not expect a different tax evasion strategy between both Self-Disclosure Treatments in any period.

However, we have demonstrated that we indeed find an increased tax evasion behavior in the Self-Disclosure without Penalty Treatment. Compared to the No Self-Disclosure Treatment, subjects in the Self-Disclosure with Penalty Treatment only increase tax evasion in those periods where we expect risk-neutral subjects to do so. In contrast to these findings, subjects in the Self-Disclosure without Penalty Treatment also increase tax evasion beyond these two periods so that we find significant decreased tax compliance after the introduction of uncertain audit probabilities. Therefore we want to analyze what drives participants to increasingly evade taxes in the Self-Disclosure without Penalty Treatment compared to the Self-Disclosure with Penalty Treatment or, vice versa, what limits subjects' tax evasion behavior in the Self-Disclosure with Penalty Treatment.

As both Self-Disclosure Treatments only differ in the penalty rate or the existence of a penalty at all, we find two factors that might be influenced by the penalty rate. First, the existence of a possible penalty increases the complexity of the tax evasion decision compared to an environment in which the subject

[33] We refrain from analyzing model A32 as the model fit Prob > chi2 is too high (0.9997).

is granted a self- disclosure option with impunity.[34] Numerous studies already have shown that complexity (cognitive load) decreases risk-taking. Hence, complexity should also reduce tax evasion. Second, if the existence of a penalty on the self-disclosed evaded taxes makes tax evasion appear more illegal than under the self-disclosure option without penalty, the illegalization of the self-disclosure opportunity can lead to an illegalization of tax evasion itself, thus reducing tax evasion behavior.

We start by analyzing whether subjects perceive the decisions' complexity of the Self-Disclosure without Penalty Treatment and of the Self-Disclosure with Penalty Treatment differently. In the post-experimental questionnaire we asked participants: "How complex do you perceive the decision periods?" On an 11-point scale subjects could rate their opinion, whereby 0 presents "not complex at all" and 10 "very complex".[35] Whereas subjects in the Self-Disclosure without Penalty Treatment rate COMPLEX DECISION with 1.792 on average, subjects in the Self-Disclosure with Penalty Treatment rate COMPLEX DECISION with 2.655. This difference is slightly significant with a p-value of 0.0641 (2-sided Mann-Whitney U test). Hence, as predicted subjects in the Self-Disclosure with Penalty Treatment face a higher decision complexity due to the penalty than subjects in the Self-Disclosure without Penalty Treatment. Besides the analysis, we also test the time that subjects need to make their tax evasion decision. Whereas subjects in the Self-Disclosure with Penalty Treatment need about 11.05 seconds to make their decision, subjects in the Self-Disclosure without Penalty Treatment need about 16.122 seconds. This difference is not significant with a p-value of 0.9879 (2-sided Mann-Whitney U test).[36] Hence, we find some evidence that the decisions complexity differs between the treatments but may not prove that this is mirrored in the decision time.

Second, we analyze whether the existence of a penalty within the self-disclosure option may lead to an illegalization of the option itself which might reduce tax evasion. For this we analyze the emotions that were perceived by the subjects in the course of the experiment and rated on an 11-point scale in the post-experimental questionnaire. For the emotions JOY, ANGER and ANXIETY we do not find any significant differences between the Self-Disclosure without Penalty and the Self-Disclosure with Penalty Treatment.[37] However, subjects state their feelings of guilt differently in both treatments. Whereas subjects in the Self-Disclosure without Penalty Treatment indicate GUILT with 0.833, subject in the Self-Disclosure with Penalty Treatment denote GUILT to be 1.828. Although both values are relatively low, the difference is significant with $p = 0.0168$ (2-sided Mann-Whitney U test). Thus, even though tax

[34] See e.g., Dohmen et al. (2010).

[35] We have already introduced the respective variable in section **Fehler! Verweisquelle konnte nicht gefunden werden.** for the multivariate analyses.

[36] The average time in the Self-Disclosure without Penalty treatment is mainly driven by one participant who needs a lot more time for making the decision (139.417 seconds on average) and thus skews the results. Dropping these observations decreases the average time to 10.76 seconds but still reveals no significant difference ($p = 0.4333$).

[37] For the emotion JOY subjects indicate their feeling with 5.29 (5.38; p-value = 0.9427), for ANGER with 3.67 (3.86; p-value = 0.8071) and for ANXIETY with 1.71 (1.97; p-value = 0.7967) in the Self-Disclosure without Penalty treatment (Self-Disclosure with Penalty treatment; p-value calculated by 2-sided Mann-Whitney U tests).

evasion rates are significantly lower in the Self-Disclosure with Penalty Treatment, subjects feel guiltier in this treatment. This result is rather astonishing but indicates that the introduction of a penalty into a self-disclosure option leads to an illegalization of the option itself and criminalizes the act of tax evasion. Summarizing, we find evidence for both explanations to limited increased tax evasion in the Self-Disclosure with Penalty Treatment: an increased complexity and an illegalization of the self-disclosure option, both due to the penalty rate.

5.6 Conclusions

We run a standard tax evasion game with repetition and introduce a self-disclosure opportunity for two of three treatment groups (between-subject design). While one self-disclosure treatment entails a penalty payment on the revealed evaded taxes, the other self-disclosure treatment grants impunity when using the option. Additionally, we spilt the experiment into two phases: The first five periods are run with certain audit probabilities, whereas in the last seven periods two possible audit probabilities are presented of which one is randomly drawn after the tax evasion decision. After each tax evasion decision (and after the decisive audit probability is drawn), subjects in the self-disclosure treatments have the opportunity to repay the evaded taxes.

In our study we examine whether we find evidence for two major points of criticism that are often raised: First, we test whether the self-disclosure option erodes tax morale, thus causing a crowding out of formerly intrinsic motivated subjects that leads to increasing tax evasion (hypothesis 1). Second, we test whether subjects use the self-disclosure option strategically under uncertain audit probabilities which also increases tax evasion (hypothesis 2). In a last step, we test whether small penalties that accompany the self-disclosure option are sufficient to limit the tax evasion associated with the option (hypothesis 3).

We find that subjects commonly behave as predicted assuming risk-neutral and income maximizing individuals. They show an equal tax evasion strategy independent on whether they are offered a self-disclosure option or not if they are confronted with certain audit probabilities. As we do not find any tax evasion differences under certain audit probabilities, we do not find any evidence for the self-disclosure option to cause a crowding out of tax morale and may not confirm our first hypothesis. Additional analyses reveal that we are not able to find differences in tax morale, neither self-reported nor shown through tax evasion behavior. However, under uncertain audit probabilities subjects who are offered the self-disclosure options evade more taxes in those periods where a divergent tax evasion strategy for the treatments is predicted. We even find that subjects who are offered the self-disclosure option with impunity increase tax evasion significantly beyond the expected periods. Thus we may confirm the second hypothesis that subjects use the option strategically and integrate it into their tax evasion decision. In this context, we find that the self-disclosure option is mainly used strategically. Thus, subjects use the opportunity above all in periods with uncertain audit probabilities when the high audit probability is drawn. Finally, we are able to show that if the self-disclosure option is linked with a

penalty payment, the increase in tax evasion can be limited (hypothesis 3). We find evidence that this is driven by the illegalization of tax evasion due to the penalty when using the self-disclosure option and/or by an increase of the tax evasion decision's complexity when taking a possible penalty payment into account.

Our results are in line with former research on the impact of tax amnesties and voluntary disclosure on tax compliance (see e.g., Alm et al., 1990; Andreoni, 1991; Langenmayr, 2017) who show that tax amnesties and voluntary disclosure opportunities decrease tax compliance. We find that this decrease is not caused by a crowding out of formerly morale and honest taxpayer, but by the strategical integration of the self-disclosure option into the tax evasion decision. Although we are able to confirm some concerns related to the voluntary self-disclosure we demonstrate that even rather small penalty payments are sufficient to limit tax evasion associated with the option. They prevent subjects from extensively increasing tax evasion so that the self-disclosure option can be a useful tool to offer subjects a way back to tax honesty and can generate revenues without bearing a considerable fiscal risk if the pre-self-disclosure tax evasion level is high (Andreoni, 1991), if it is linked to stricter tax enforcement policies (Stella, 1991) or if paying taxes is adopted as social norm (Alm and Beck, 1990). Moreover, Langenmayr (2017) finds that administrative costs decrease with the introduction of a self-disclosure option. Thus, while keeping the compliance rate relatively constant, the self-disclosure option with a small penalty is able to decrease administrative costs. We therefore argue that the self-disclosure opportunity is a suitable way to generate additional revenues without decreasing tax compliance significantly if it contains penalty payments.

Nevertheless, our study is only a first step in analyzing the self-disclosure option experimentally and may be extended in some respects. First, we refrain from modelling a tax evasion setting which comprises more than one period so that the tax evasion and self-disclosure decisions are only relevant for the respective period. Although this setting is rather theoretical and does not completely depict real voluntary self-disclosure mechanism it is a useful tool to get a first impression on the voluntary disclosure's impact and effect. However, further research could integrate a model which comprises more periods if the subject is audited or uses the voluntary disclosure opportunity, but has to carefully analyze the data by accounting for the periods' interdependency. Another extension of the presented project is to model the setting as public good game and analyze the impact on morale constraints, social norms and learning of tax evasion strategies. So far, public good experiments only analyze tax amnesties (Alm et al., 1990; Torgler and Schaltegger, 2005; Rechberger et al., 2010). Another extension of our study can be the integration of the tax authority as rational player with limited audit capacities and a tax revenue maximizing strategy. In that case, audit probabilities are strategically chosen and the impact of the self-disclosure opportunity on tax revenues in the short- and in the long run could be analyzed. However, also in this model the interdependence of taxpayer and tax authority has to be taken into consideration and analyzed respectively.

5.7 Appendix A: Experimental Instructions

Below are the experiment's instructions. The experiment was conducted in German, hence, everything shown below was translated into English.

Welcome to our experiment

Thank you very much for attending this experiment. The experiment will last about one hour. How much you earn depends on your decisions and on chance. This instruction illustrates how your decisions influence how much money you earn in the experiment. Therefore, read the following paragraphs carefully. In the following we give you a short overview on the experiment's procedure.

Before the experiment starts we want to raise some important points.

- You are not allowed to communicate with other participants or leave your place during the whole experiment. Please also remain on your workplace with your eyes.
- Please turn off your mobile phone and put it into your back.
- Please read the instructions closely and carefully.
- It is important that you have understood the instructions. Thus, please do not hesitate to ask questions. If you have questions, please raise your hand. We will come to you to answer your questions. Please do not ask your questions aloud.
- You may highlight or write on the instructions.
- You may use the pen in front of you.
- Please do not take the instructions home but hand them back to the experimenter at the end of the experiment.
- The program that runs the experiment – this is the grey part on your screen – may not be closed. Please also do not open any other programs as this may cause a termination of the whole experiment.
- Please note, that there may occur latency in the experiment as different participants are differently fast. Thus, please do not wonder if you are asked to wait several minutes.

You find the experiment's instructions on the next pages.

Decision periods

The experiment consists of 12 periods in total which are independently of one another. At the end of the experiment one period is randomly drawn. This period's decisions determine your payoff of the experiment.

In each of the 12 periods you are assigned an income which amounts to 1,000 cents. Out of these 1,000 cents you must pay a contribution of 40% of your income (i.e., 400 cents). The level of contribution (40%) indeed is compulsory, but you may ultimately decide on your own how much you want to contribute as your anonymity is granted. All contributions which are made in this experiment are used for further research at the Leibniz University of Hannover.

Your task is to decide in every period how much contribution you want to pay. With a certain probability your declaration is checked. These probabilities are displayed to you at the beginning of each period before you have to make your decision. These probabilities may differ from period to period and take values between 0% and 80%.

If there is a check and your contribution does not meet 40% (400 cents) you have to pay the discrepancy. Additionally, you have to pay a penalty which amounts to 100% of the missing contribution. This means the whole remargin amounts to twice the discrepancy contribution. If you have paid all the demanded contribution in the period checked, no penalty is raised and you do not have to remargin.

[Voluntary Self-Disclosure with Penalty Treatment only: In every period you have the opportunity to obtain a partial amnesty. After your decision but before you are eventually checked you can decide on your own whether you want to use this opportunity. In order to receive the partial amnesty you have to fully pay the discrepancy. In return, the additional penalty only amounts to 10% (instead of 100%) of the missing contribution.]

[Voluntary Self-Disclosure without Penalty Treatment only: In every period you have the opportunity to obtain an amnesty. After your decision but before you are eventually checked you can decide on your own whether you want to use this opportunity. In order to receive the amnesty you have to fully pay the discrepancy. In return, there is no additional penalty (instead of 100% of the missing contribution).]

After every period you receive the information whether you were checked or not. Furthermore, you get an overview of all important values as well as your period's payment.

Filling in a questionnaire

Subsequent to the last decision period we would like to ask you to fill in a questionnaire which will start automatically after the experiment on your screen once all participants have made the last experiment's decision. The questionnaire's data is solely required for research purposes and are really important to us. Under no circumstances we pass the data to third parties! Moreover, we may not use the data to conclude on you so that your anonymity is always granted.

Experiment's Payment

After you have made your decisions in all 12 periods, *one* period is randomly drawn by the computer at the end of the experiment and displayed to you on the screen. This period's result which resulted from your decision in the respective period is converted to euro. Each Participant additionally receives a show-up free amounting to 3 euro for taking part in the experiment. The resulting entire payment is paid to you cash subsequent to the experiment.

Comprehension Test

Before the initial experiment starts you are asked to answer some comprehension questions regarding the experiment in order to ensure that you fully understand your tasks. However, they are not relevant for payoff. If you are not able to answer a question please raise your hand. We come and help you.

The experiment starts when all participants are ready.

5.8 Appendix B: Screenshots of the Experiment

Below are screenshots from the experiment. The experiment was conducted in German, hence, everything shown below was translated into English.

Please note, that the following first ten screenshots are only valid for the experiment's first five

Figure 5.3: Tax Evasion Decision Stage in the First Five Periods

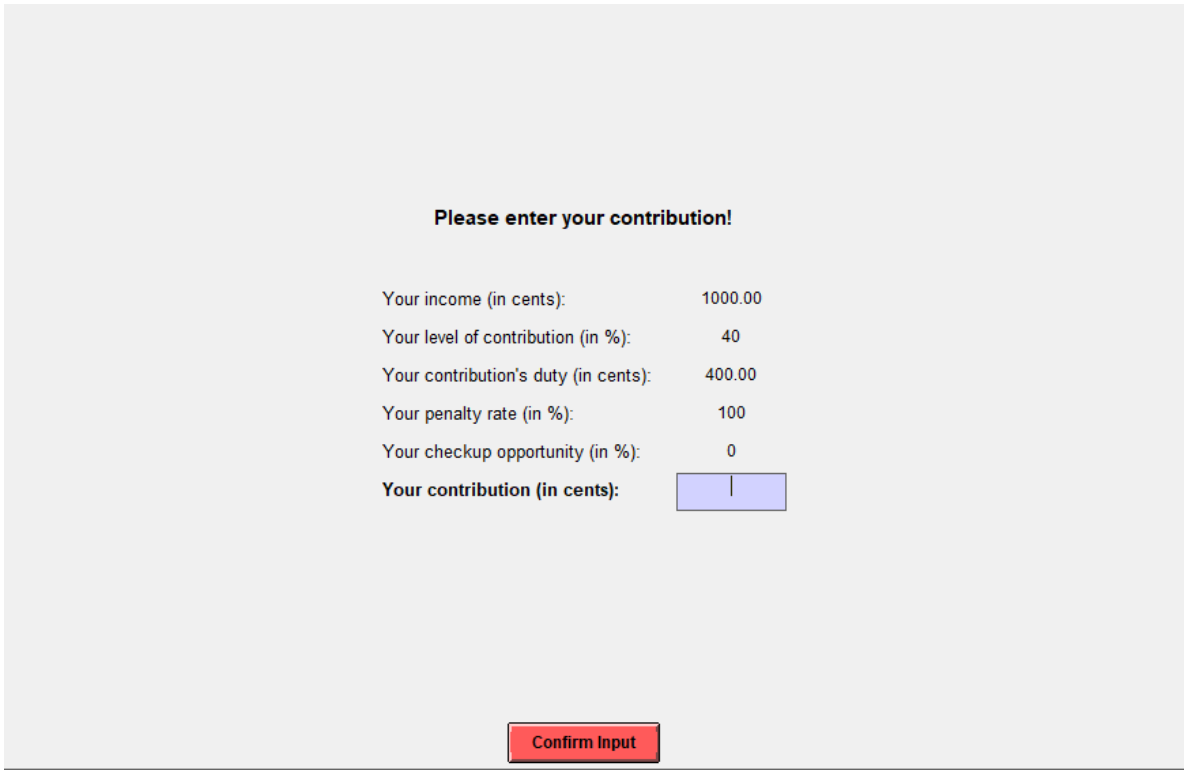


Figure 5.4: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the First Five Periods

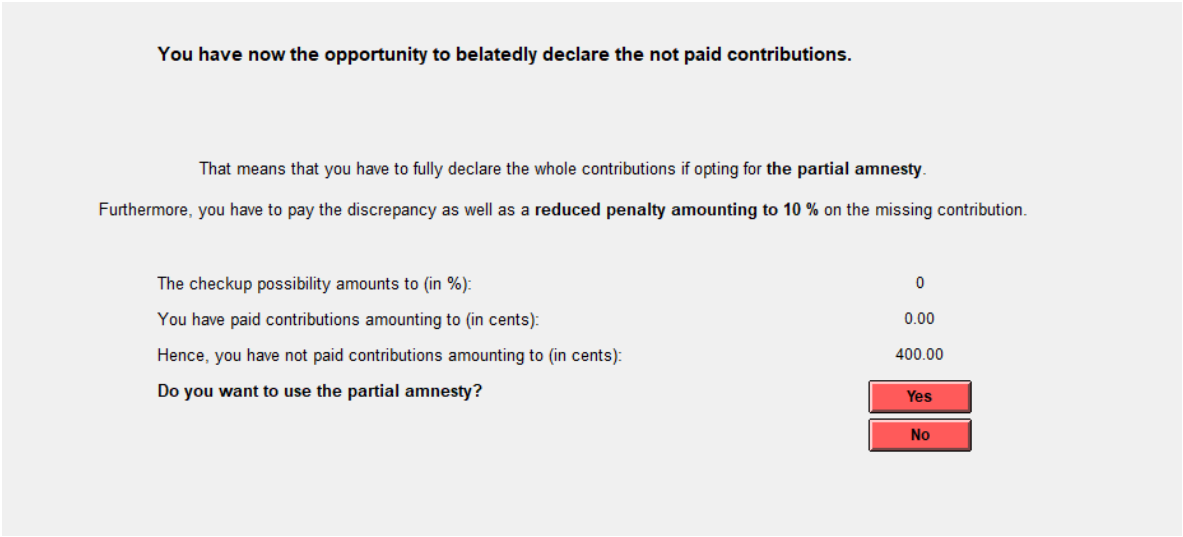


Figure 5.5: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the First Five Periods

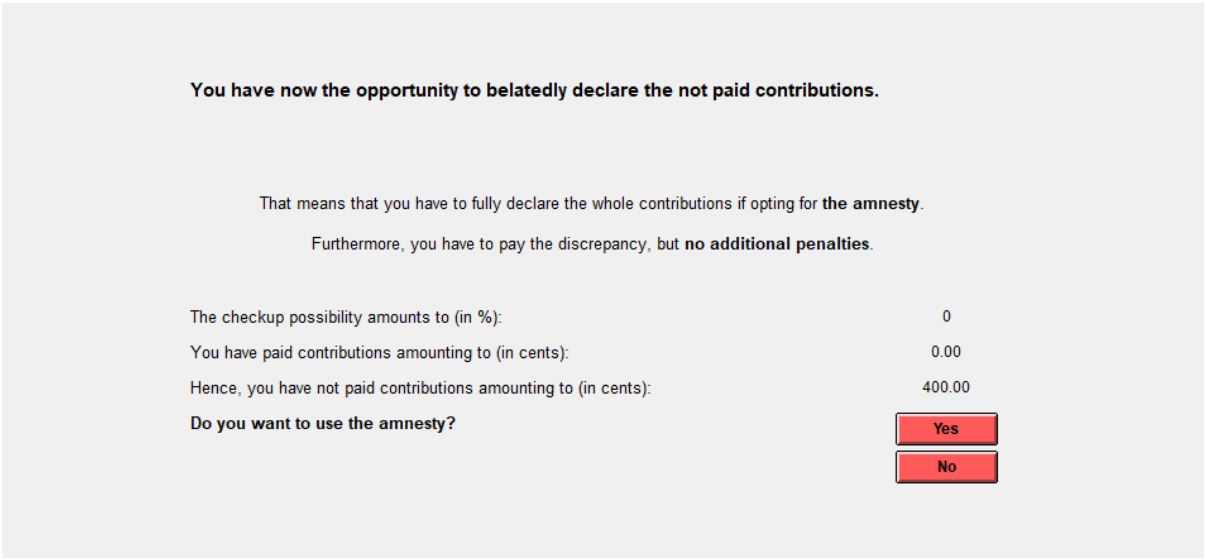


Figure 5.6: Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the First Five Periods if all Taxes are already Declared Truthfully in the First Five Periods

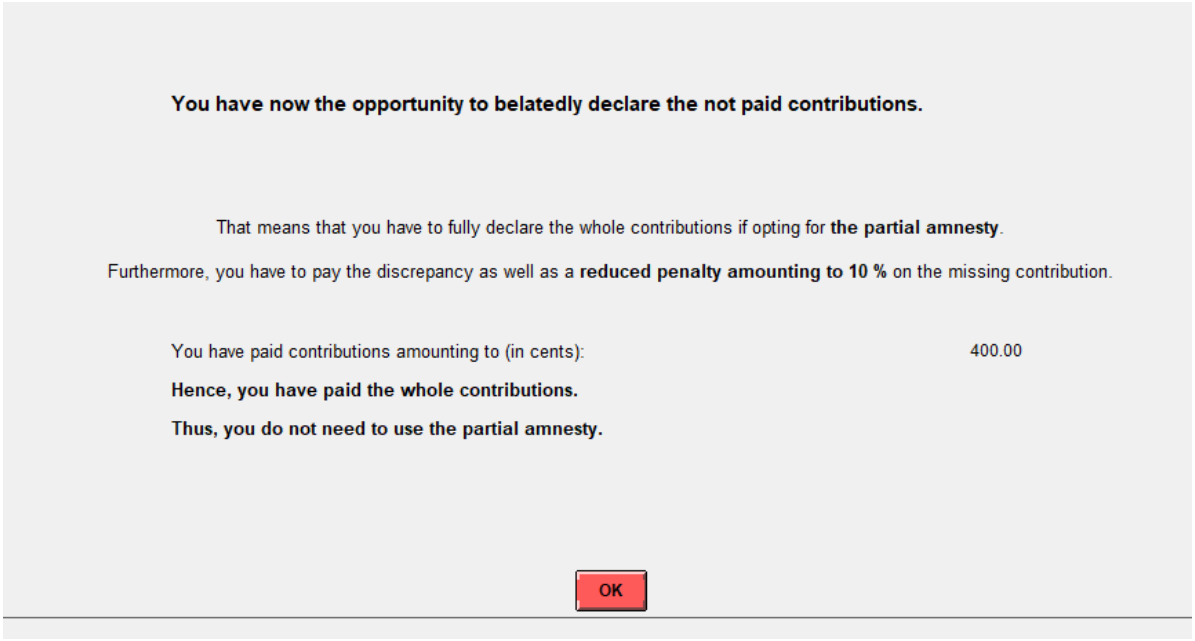


Figure 5.7: Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the First Five Periods if all Taxes are already Declared Truthfully in the First Five Periods

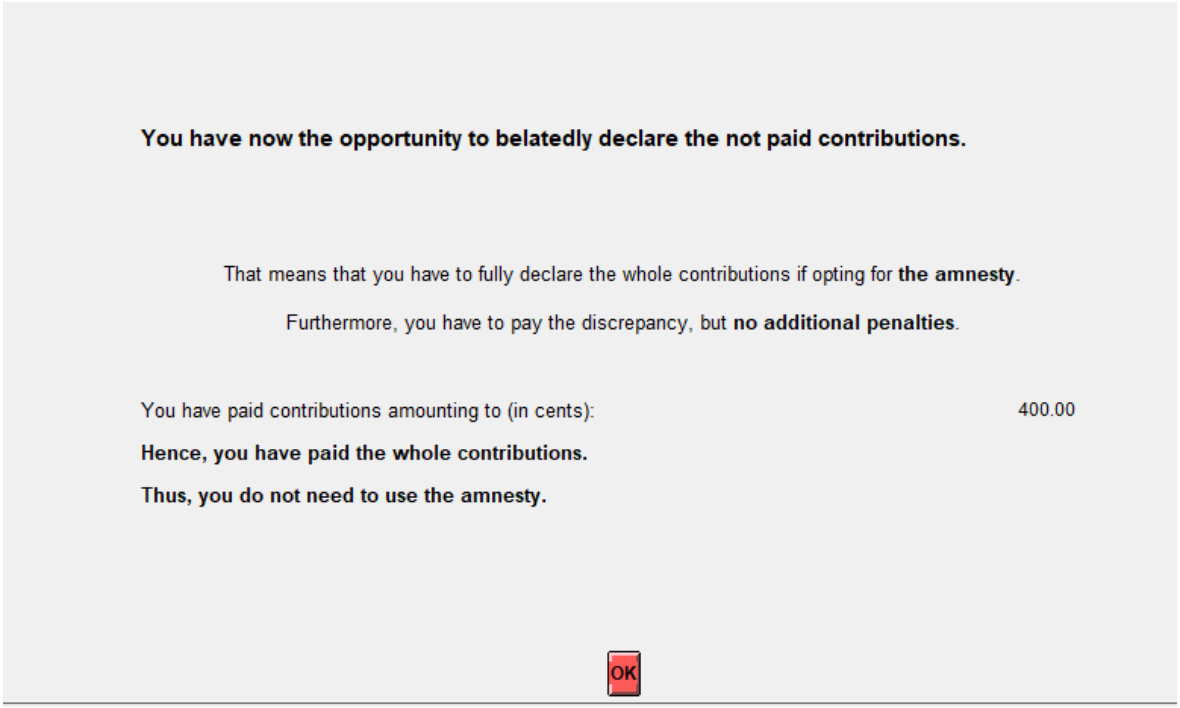
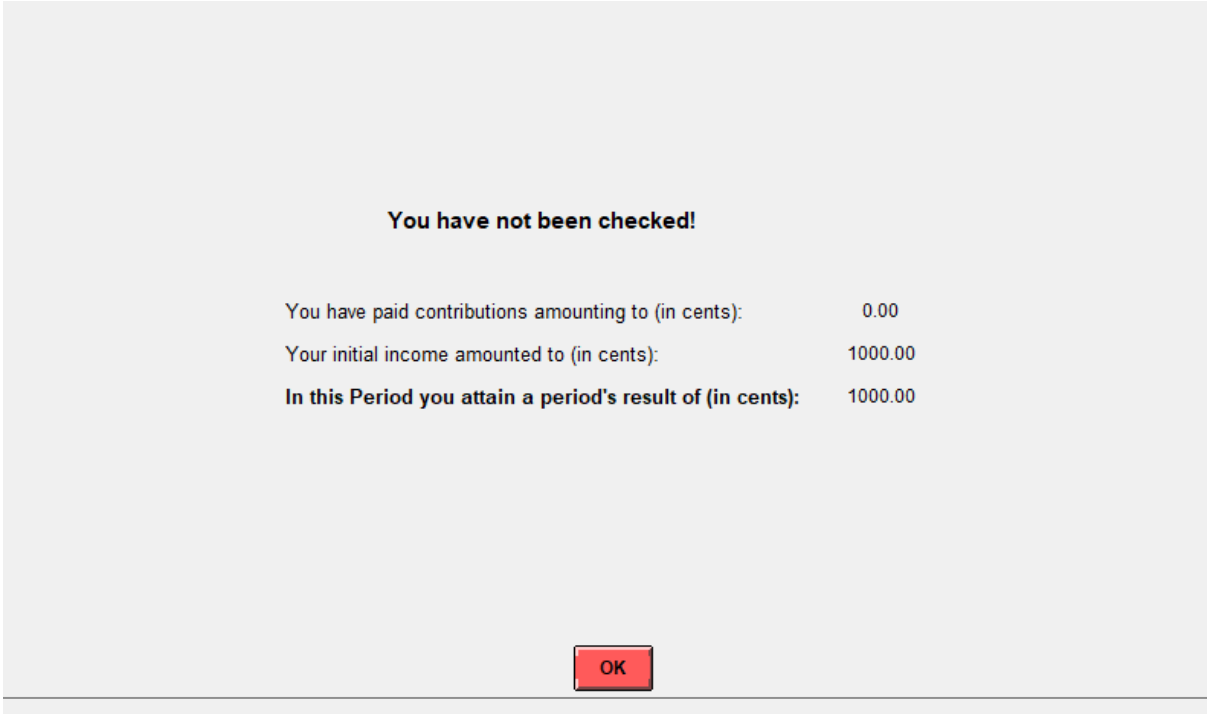


Figure 5.8: Information Stage in the First Five Periods if the Subject is not Audited



Note: This screen is valid for all subjects (in the Self-Disclosure Treatments) who are not audited (and did not use the self-disclosure option).

Figure 5.9: Information Stage in the First Five Periods if the Subject is in the Self-Disclosure with Penalty Treatment and Uses the Self-Disclosure Opportunity

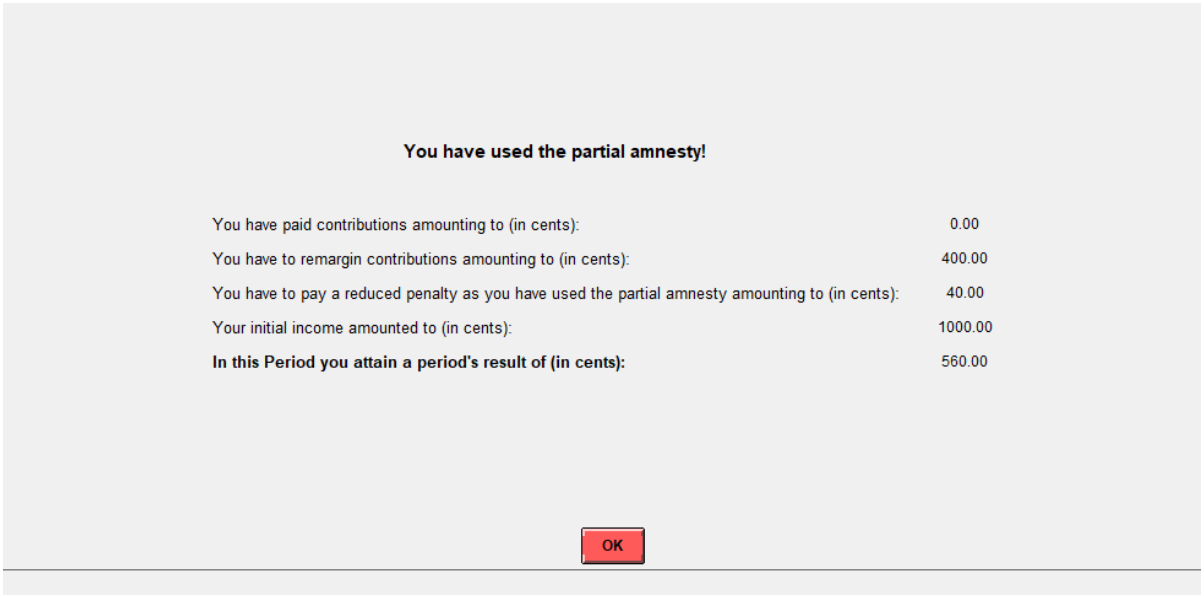


Figure 5.10: Information Stage in the First Five Periods if the Subject is in the Self-Disclosure without Penalty Treatment and Uses the Self-Disclosure Opportunity

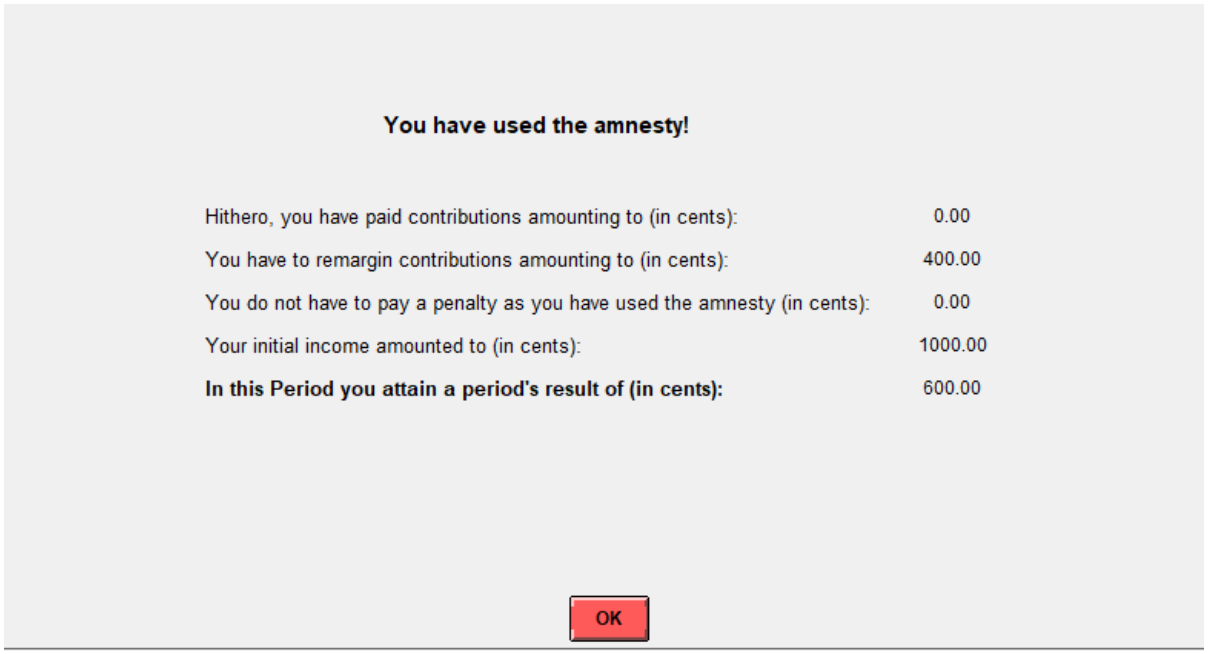
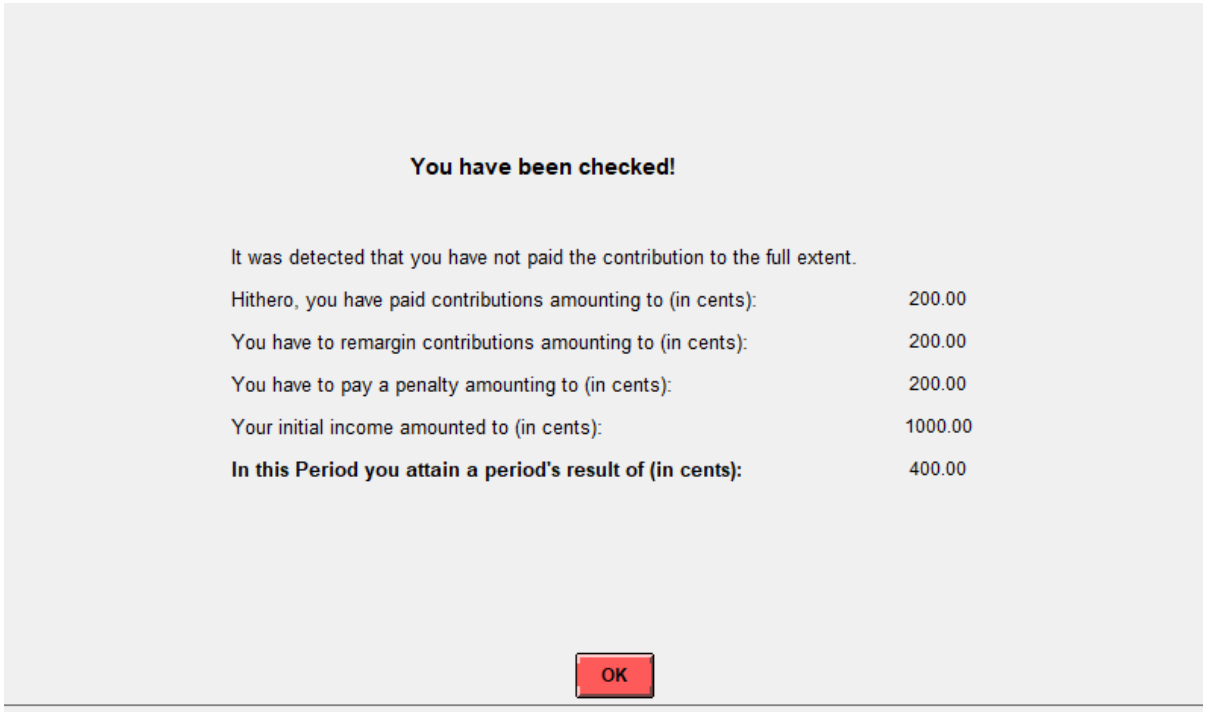
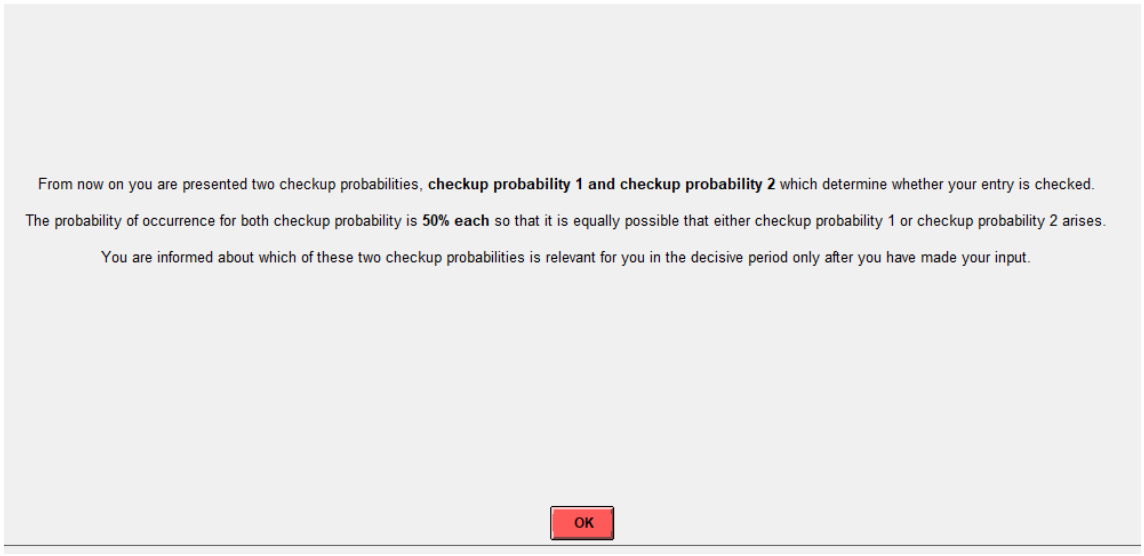


Figure 5.11: Information Stage in the First Five Periods if the Subject is Audited



Note: This screen is valid for all subjects (in the Self-Disclosure Treatments) who are audited (and did not use the self-disclosure option).

Figure 5.12: Information on the Introduction of Two Audit Probabilities



Please note, that the following screenshots are only valid for the experiment's last five periods.

Figure 5.13: Tax Evasion Decision Stage in the Last Seven Periods

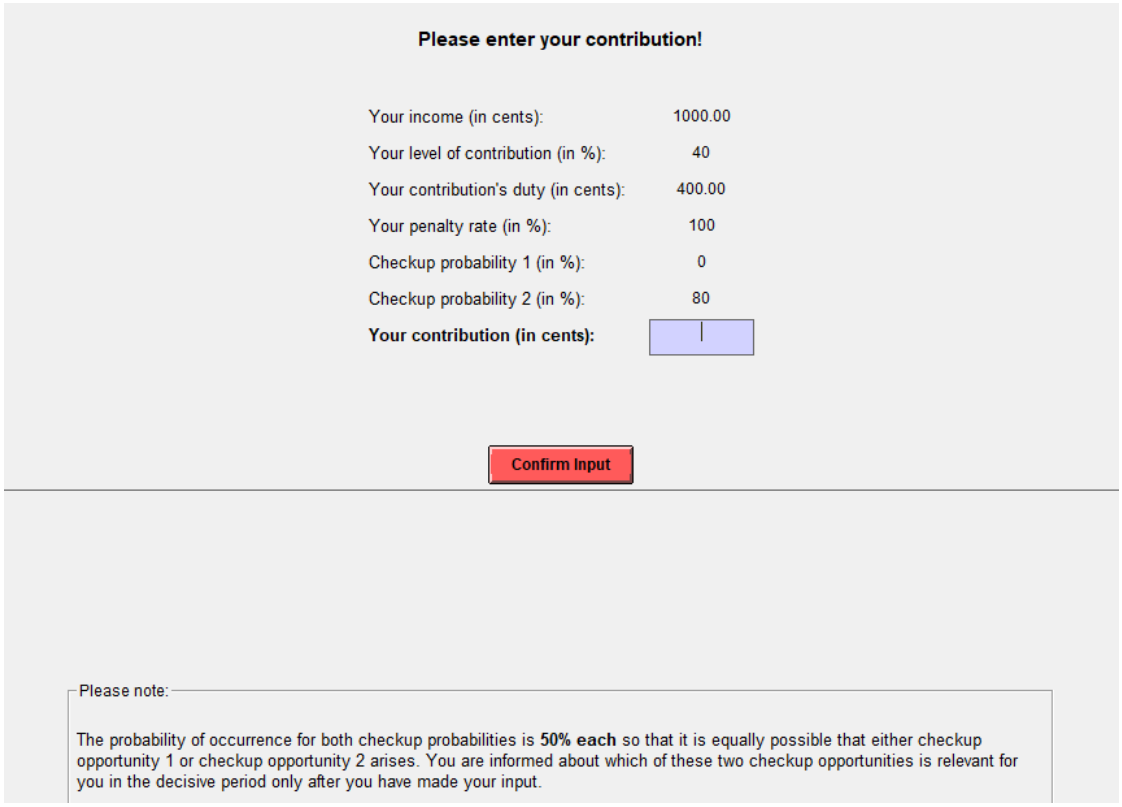


Figure 5.14: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the Last Seven Periods

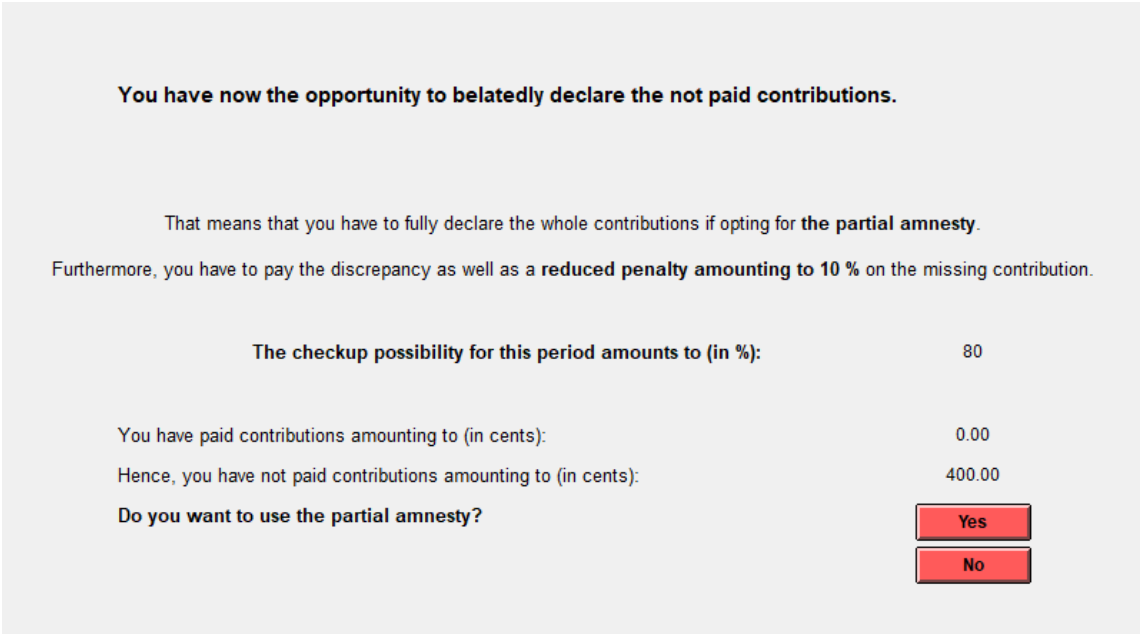


Figure 5.15: Decision on Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the Last Seven Periods

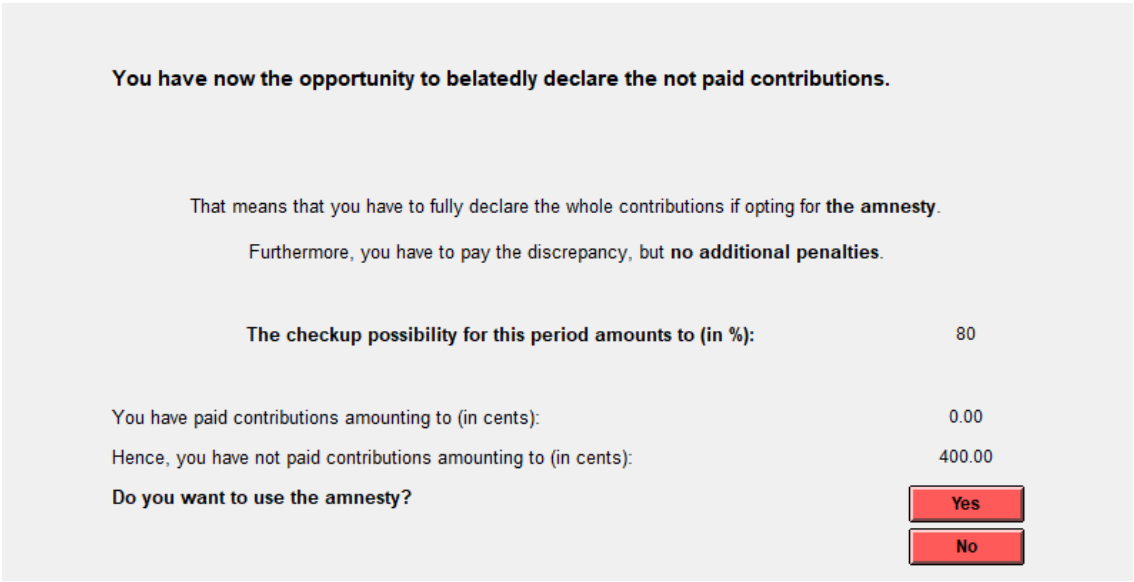


Figure 5.16: Voluntary Self-Disclosure Stage in the Self-Disclosure with Penalty Treatment in the Last Seven Periods if all Taxes are already Declared Truthfully

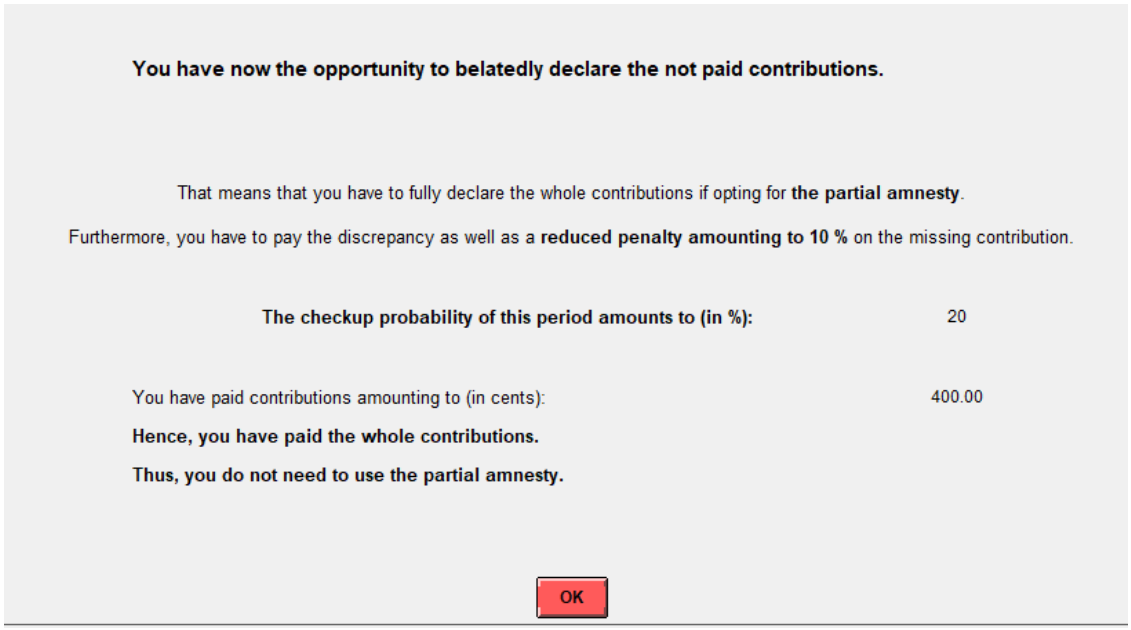


Figure 5.17: Voluntary Self-Disclosure Stage in the Self-Disclosure without Penalty Treatment in the Last Seven Periods if all Taxes are already Declared Truthfully

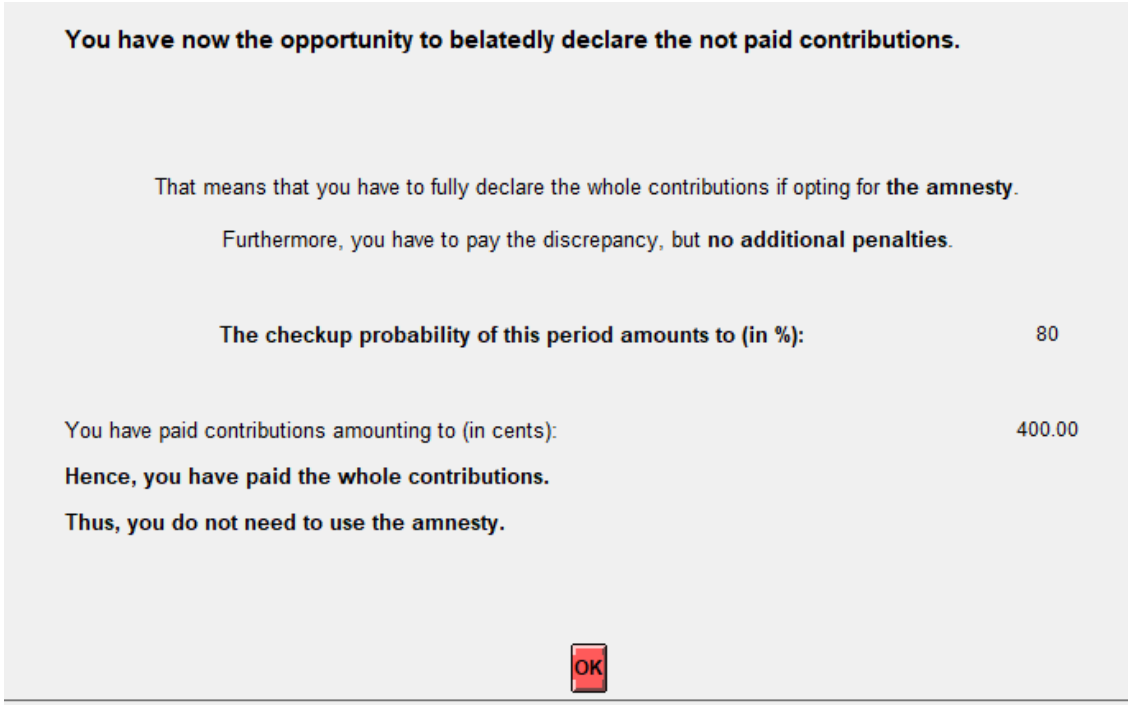


Figure 5.18: Information Stage in the Last Seven Periods if the Subject is in the Self-Disclosure with Penalty Treatment and uses the Self-Disclosure Opportunity

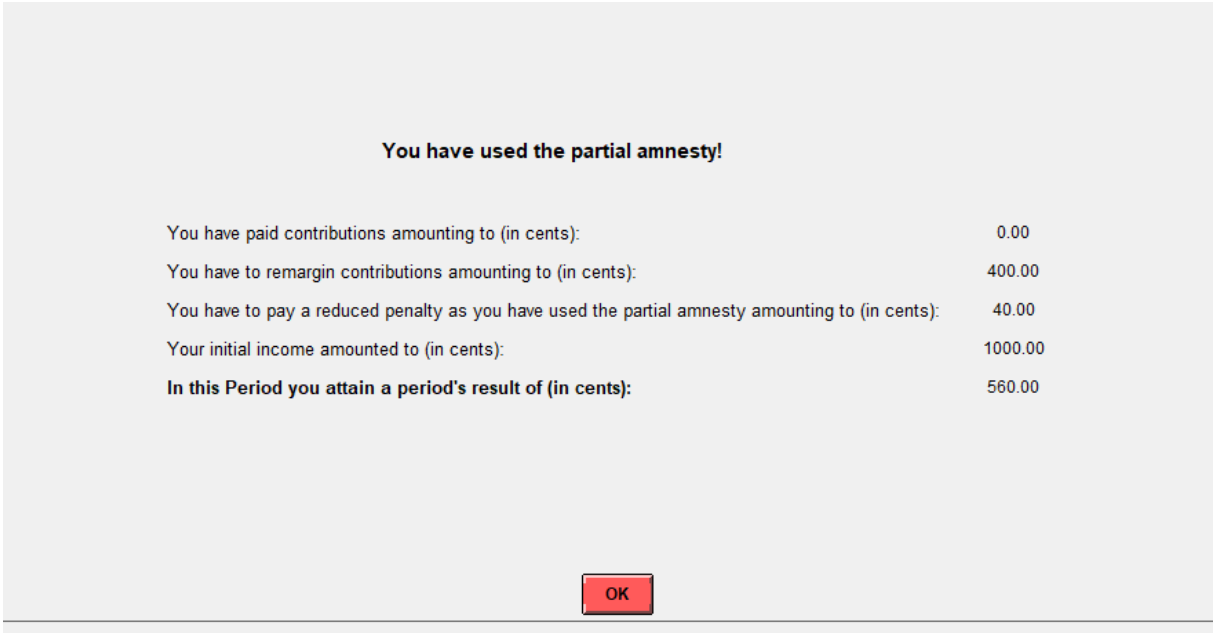


Figure 5.19: Information Stage in the Last Seven Periods if the Subject is in the Self-Disclosure without Penalty Treatment and uses the Self-Disclosure Opportunity

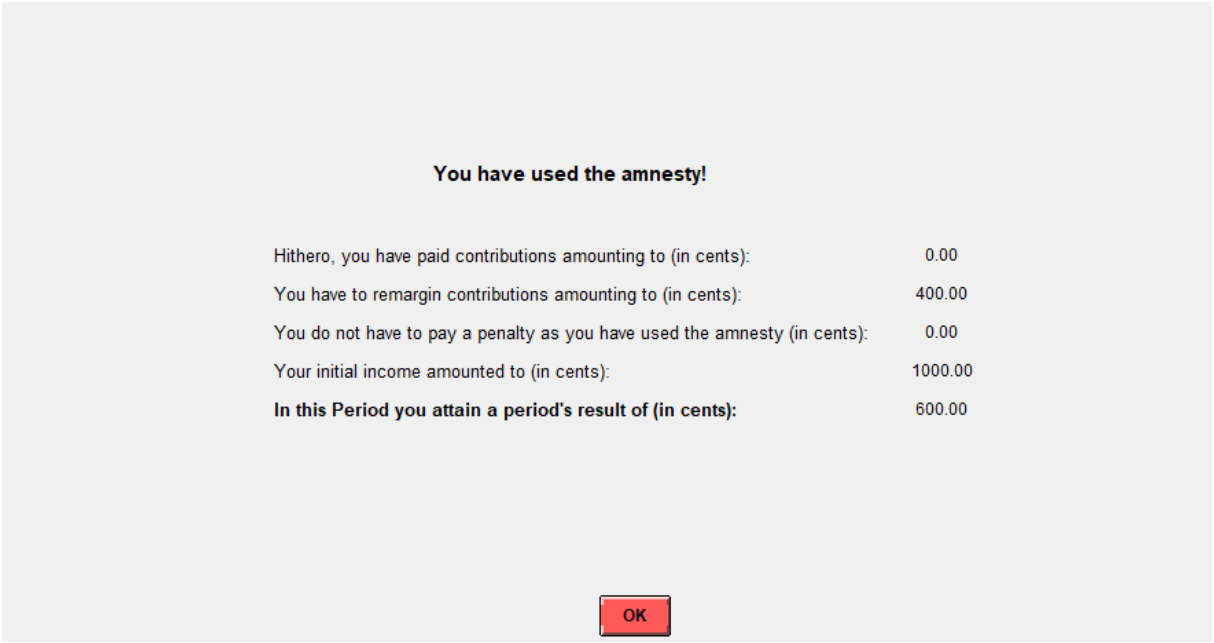
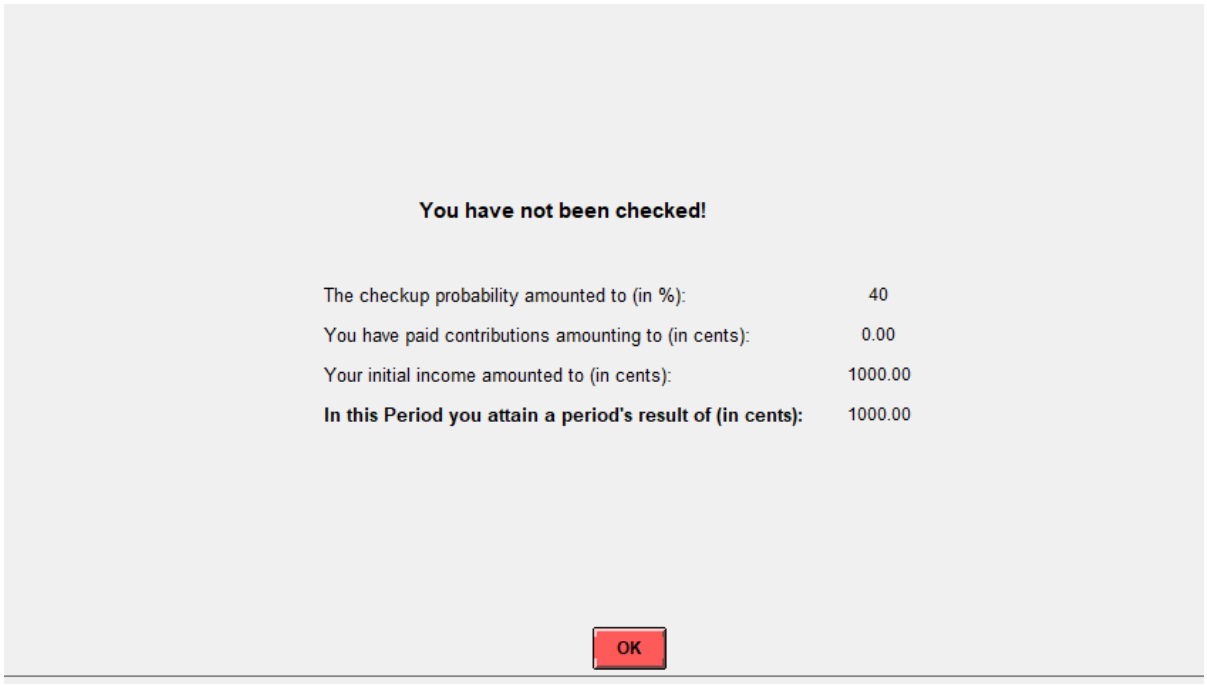
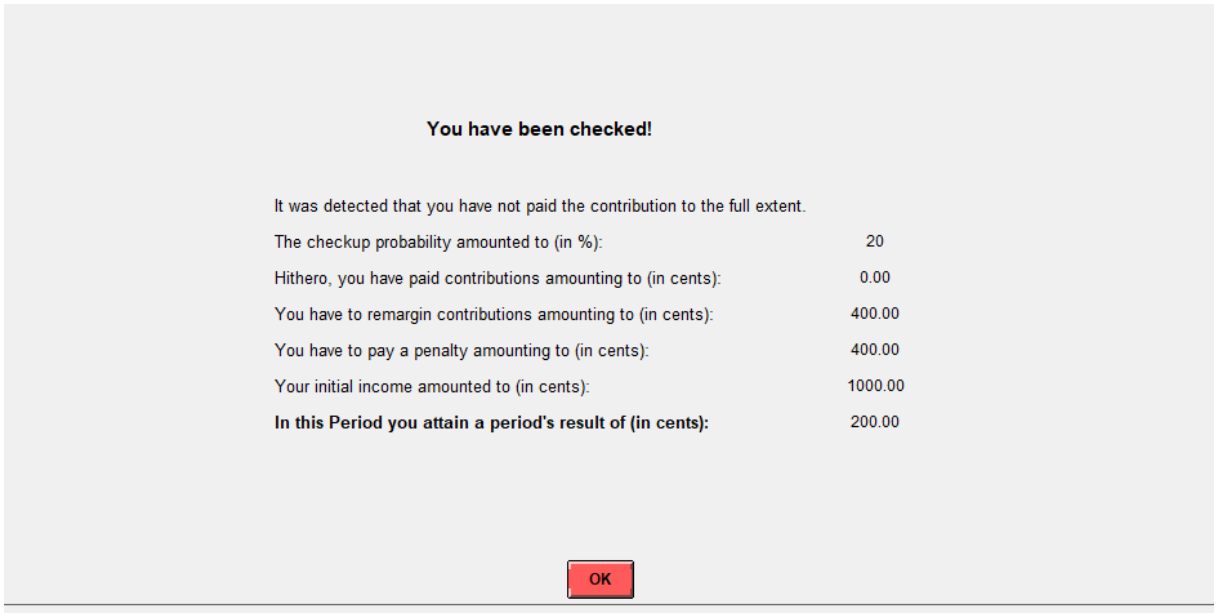


Figure 5.20: Information Stage in the Last Seven Periods if the Subject is not Audited



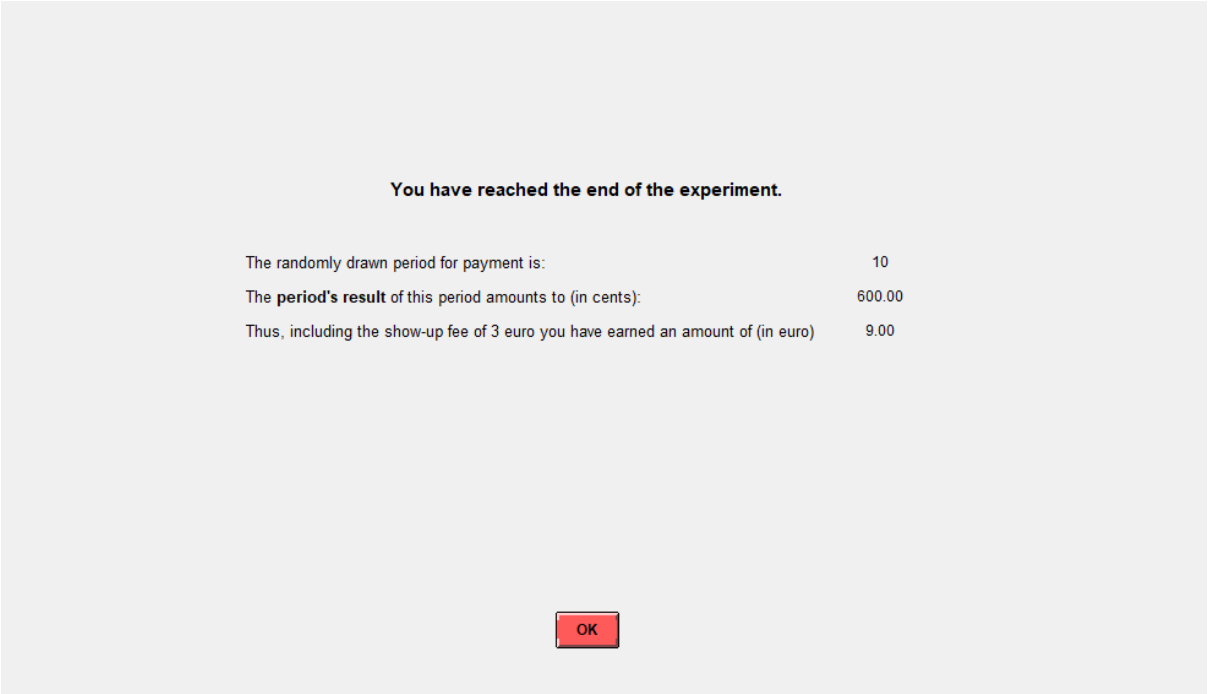
Note: This screen is valid for all subjects (in the Self-Disclosure Treatments) who are not audited (and did not use the self-disclosure option).

Figure 5.21: Information Stage in the Last Seven Periods if the Subject is Audited



Note: This screen is valid for all subjects (in the Self-Disclosure Treatments) who are audited (and did not use the self-disclosure option).

Figure 5.22: Information Stage on Experiment's Payoff



5.9 Appendix C: Questionnaire

Below is the questionnaire from the experiment. The questionnaire was also conducted with z-tree and in German, hence, everything shown below was translated into English.

How old are you (in years)? ☐

Are you female or male?

☐ Female ☐ Male

Which faculty are you enrolled for?

- ☐ Architecture and landscape
- ☐ Construction engineering and geodesy
- ☐ Electrical engineering and computer science
- ☐ Law
- ☐ Mechanical engineering
- ☐ Mathematics and physics
- ☐ Natural sciences
- ☐ Philosophy,
- ☐ Business Management and Economics
- ☐ Others
- ☐ I am not a student

What qualification are you aiming at right now?

- ☐ Bachelor
- ☐ Master
- ☐ Diploma
- ☐ State examination
- ☐ Others

Which academic semester are you in? ☐

How would you rate yourself: Are you generally risk-taking or do you try to avoid risk?

Not risk-loving at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very risk-loving

Do you ever have submitted a tax return?

- ☐ Yes – once
- ☐ Yes – twice
- ☐ Yes – more than twice
- ☐ No
- ☐ I do not know

Please judge the following statements.

	You are not allowed to do this on any terms										It is all right in any case
Evading taxes if you have the possibility to do so.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

	On no account										On any account
If somebody does something good to you, are you content to reciprocate even if it was not agreed upon beforehand?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
If someone treats you badly, do you also treat this person badly?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Very unfair										very fair
How do you rate the contribution and control system which was applied within the experiment?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Not complex at all										Very complex
How complex do you perceive the decision periods (fiscal decisions)?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please state, which emotions you have perceived in the course of the experiment.

	It is not true at all										It is absolutely true
I have perceived joy.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived anger.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived anxiety.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
I have perceived guilt.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

What is your monthly disposable income (after rent)?

How often do you pray per week?

- ☐ Never
- ☐ 1 – 2 times
- ☐ 3 – 5 times
- ☐ Daily
- ☐ Several times a day

5.10 Appendix D: Derivations

D1 Determination of the Optimal Tax Evasion Strategy in the No Self-Disclosure Treatment given a Certain Audit Probability

We derive the critical audit probability which changes tax evasion strategies by comparing the expected payoff of the tax evasion decision with the payoff of full tax compliance in the No Self-Disclosure Treatment under certain audit probabilities.

$$\pi_C = \mathrm{E}\left[\pi_{NoSD}^{Certain}\right] \tag{D5.1}$$

$$\mathrm{E} - \mathrm{T} = (1 - \mathrm{p})(\mathrm{E} - \mathrm{D}) + \mathrm{p}[\mathrm{E} - \mathrm{D} - (T - D)(1 + F)] \tag{D5.2}$$

$$\Leftrightarrow \mathrm{D} - \mathrm{T} = \mathrm{p}(\mathrm{D} - \mathrm{T})(1 + F)$$

$$\Leftrightarrow \mathrm{p}^* = \frac{1}{1 + F} \tag{5.5}$$

Another option to derive the critical audit probability is to differentiate the expected payoff with respect to the declared taxes D.

$$\mathrm{E}[\pi_{NoSD}^{Certain}] = (1 - \mathrm{p})(\mathrm{E} - \mathrm{D}) + \mathrm{p}[\mathrm{E} - \mathrm{D} - (T - D)(1 + F)] \tag{5.1}$$

$$\frac{\mathrm{dE}[\pi_{NoSD}^{Certain}]}{dD} = -(1 - \mathrm{p}) + \mathrm{p}[-1 + (1 + F)]^{!};=0 \tag{D5.3}$$

$$\Leftrightarrow -1 + \mathrm{p}(1 + F)^{!};=0$$

$$\Leftrightarrow \mathrm{p}^* = \frac{1}{1 + F} \tag{5.5}$$

Up to an audit probability of $\mathrm{p} = \frac{1}{1+F}$ complete tax evasion is always rational for risk-neutral subjects. For all audit probabilities above this term full compliance maximizes the expected payoff.

D2 Determination of the Optimal Tax Evasion Strategy in the No-Self Disclosure Treatment given Uncertain Audit Probabilities

We derive the critical audit probability which changes tax evasion strategies by comparing the expected payoff of the tax evasion decision with the payoff of full tax compliance in the Self-Disclosure Treatment under uncertain audit probabilities.

$$\pi_C = \mathrm{E}\left[\pi_{NoSD}^{Uncertain}\right] \tag{D5.4}$$

$$\mathrm{E} - \mathrm{T} = \theta[p_H\pi_A + (1-p_H)\pi_{NA}] + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}] \tag{D5.5}$$

$$\Leftrightarrow \mathrm{D} - \mathrm{T} = (\theta p_H - \theta p_L + p_L)(\mathrm{D} - \mathrm{T})(1+F)$$

$$\Leftrightarrow \frac{1}{1+F} = \theta(p_H - p_L) + p_L \tag{D5.6}$$

Inserting the experiment's parameters with *θ=50%* and *F=100%* we derive the following equation:

$$\Leftrightarrow \frac{1}{(1+1)} = 0.5(p_H - p_L) + p_L$$

$$\Leftrightarrow 1 = p_H + p_L \tag{D5.7}$$

$$\Leftrightarrow \frac{p_H + p_L{}^*}{\theta} = 0.5 \tag{D5.8}$$

Assuming risk-neutral subjects, subjects should evade all taxes as long as the sum of low and high audit probability is below 100%. In other words, subjects should evade all taxes as long as the expected audit probability $\frac{p_H+p_L{}^*}{\theta}$ does not exceed 50%. This is true for five of seven periods. In the periods 9 and 12, the audit probabilities amount to $p_L = 40\%$ and $p_H = 80\%$, so that risk-neutral subjects should declare their whole income truthfully.

Another option to derive the critical audit probability which changes tax evasion strategies is to derivate the expected payoff with respect to the declared taxes D.

$$\mathrm{E}\left[\pi_{NoSD}^{Uncertain}\right] = \theta[(1-p_H)\pi_{NA} + p_H\pi_A] + (1-\theta)[(1-p_L)\pi_{NA} + p_L\pi_A]. \tag{5.7}$$

$$\Leftrightarrow \mathrm{E}\left[\pi_{NoSD}^{Uncertain}\right] = \theta\left[(1-p_H)(E-D) + p_H\left(E-D-(T-D)(1+F)\right)\right]$$
$$+(1-\theta)[(1-p_L)(E-D) + p_L(E-D-(T-D)(1+F))].$$

$$\frac{\mathrm{dE}\left[\pi_{NoSD}^{Uncertain}\right]}{dD} = -\theta(1-p_H) + \theta p_H F - (1-\theta)(1-p_L) + (1-\theta)p_L F^{!};=0 \tag{D5.9}$$

$$\Leftrightarrow 1 = \theta(p_H - p_L)(1+F) + p_L(1+F)$$

$$\Leftrightarrow \frac{1}{(1+F)} = \theta(p_H - p_L) + p_L \tag{D5.6}$$

D3 Determination of the Optimal Voluntary Self-Disclosure Strategy

We derive the critical audit probability from which onwards the voluntary self-disclosure opportunity is optimal by comparing the payoff if the self-disclosure option is used (equation 5.6) and if it is not used (equation 5.1) assuming risk neutral and income maximizing subjects.

$$\pi_{SD} = \mathrm{E}\left[\pi_{NoSD}^{certain}\right] \tag{D5.10}$$

$$\begin{aligned} \mathrm{E} - \mathrm{D} - (T - D)(1 + f) &= (1 - \mathrm{p})(\mathrm{E} - \mathrm{D}) + \mathrm{p}[\mathrm{E} - \mathrm{D} - (T - D)(1 + F)]. \\ \Leftrightarrow \mathrm{E} - \mathrm{D} - (T - D)(1 + f) &= \mathrm{E} - \mathrm{D} - \mathrm{p}(T - D)(1 + F) \\ \Leftrightarrow \mathrm{p} &= \frac{1 + f}{1 + F} \end{aligned} \tag{D5.11}$$

Inserting the experiment's parameters with f = 10% (f = 0%) and F = 100% we derive a critical audit probability of 55% (50%) in the Self-Disclosure with (without) Penalty Treatment. For all periods in which the audit probability is above 55% (50%) using the self-disclosure opportunity generates a higher expected payoff than not using this opportunity.

D4 Determination of the Optimal Tax Evasion Strategy in the Self Disclosure Treatment given Uncertain Audit Probabilities

Using backward induction and the results obtained in derivation D3, we derive the critical audit probability which changes tax evasion strategies by opposing the certain payoff under complete tax compliance (equation 5.4) to the expected value of tax evasion (equation 5.8) under the assumption of risk-neutral and income maximizing subjects.

$$\pi_C = \mathrm{E}\left[\pi_{SD}^{Uncertain}\right] \tag{D5.12}$$

$$\mathrm{E} - \mathrm{T} = \theta\pi_{\mathrm{SD}} + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}]$$

$$\begin{aligned}\mathrm{E} - \mathrm{T} = {} & \theta[\mathrm{E} - \mathrm{D} - (\mathrm{T} - \mathrm{D})(1+\mathrm{f})] \\ & +(1-\theta)[p_L(\mathrm{E} - \mathrm{D} - (\mathrm{T} - \mathrm{D})(1+\mathrm{F})) + (1-p_L)(E-D)] \\ \Leftrightarrow\ & 1 = \theta(1+\mathrm{f}) + (1-\theta)p_L(1+\mathrm{F}) \\ \Leftrightarrow\ & p_L = \frac{1-\theta(1+\mathrm{f})}{(1-\theta)(1+\mathrm{F})}\end{aligned} \tag{D5.13}$$

Inserting the experiment's parameters with θ = 50%, f = 10% (f = 0%) and F = 100% we derive a critical low audit probability of 45% (50%) in the Self-Disclosure with (without) Penalty Treatment. For all periods in which the low audit probability is below 45% (50%) full tax evasion is the optimal strategy. For all low audit probabilities above 45% (50%) full tax compliance yields the highest expected payoff.

5.11 Appendix E: Tables and Figures

Table 5.4: Experiment's Decision Trees and (Expected) Payoffs for all Treatment Variations

Treatment	1st Decision	2nd Decision	(Expected) payoff		p^{crit}
No Self-Disclosure / Certainty	Tax Compliance		$\pi_C = E - T$	(5.4)	p^*=50%
	Tax Evasion		$E[\pi_{NoSD}^{certain}] = p\pi_A + (1-p)\pi_{NA}$	(5.1)	
Self-Disclosure / Certainty	Tax Compliance		$\pi_C = E - T$	(5.4)	p^*=50%
	Tax Evasion	Self-Disclosure	$\pi_{SD} = E - D - (T-D)(1+f)$	(5.6)	
		No Self-Disclosure	$E[\pi_{NoSD}^{certain}] = p\pi_A + (1-p)\pi_{NA}$	(5.1)	
No Self-Disclosure / Uncertainty	Tax Compliance		$\pi_C = E - T$	(5.4)	$\frac{p_H + p_L^*}{\theta}$ =50%
	Tax Evasion		$E[\pi_{NoSD}^{Uncertain}] = \theta[p_H\pi_A + (1-p_H)\pi_{NA}] + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}]$.	(5.7)	
Self-Disclosure / Uncertainty	Tax Compliance		$\pi_C = E - T$	(5.4)	
	Tax Evasion (θ, p_H)	Self-Disclosure	$\pi_{SD} = E - D - (T-D)(1+f)$	(5.6)	
		No Self-Disclosure	$E[\pi_{NoSD}^{pH}] = p_H\pi_A + (1-p_H)\pi_{NA}$		
	Tax Evasion ($1-\theta$, p_L)	Self-Disclosure	$\pi_{SD} = E - D - (T-D)(1+f)$	(5.6)	
		No Self-Disclosure	$E[\pi_{NoSD}^{pL}] = p_L\pi_A + (1-p_L)\pi_{NA}$		
			$E[\pi_{SD}^{Uncertain}] = \theta\pi_{SD} + (1-\theta)[p_L\pi_A + (1-p_L)\pi_{NA}]$	(5.8)	p_L^*=45%

Notes: The table presents the two decisive decisions in the in the experiment and their respective (expected) payoffs per treatment and per within variation. Thereby, the (No) Self-Disclosure Treatments are presented under certainty and uncertainty. The first decision presents the tax evasion decision which is identical for all treatments. The second decision presents the decision on a possible self-disclosure (see Figure 5.1) which is only relevant in the Self-Disclosure Treatments. The payoff in the case of (full) tax compliance (equation 5.4) and of self-disclosure (equation 5.6) is identical for all treatment variations. The expected payoff in the case of tax evasion in the No Self-Disclosure Treatment under certainty (equation 5.1) is identical with the expected payoff in the case of tax evasion in the Self-Disclosure Treatments under certainty if the self-disclosure option is not used. For the Self-Disclosure Treatments under uncertainty, using backward-induction we find that if nature draws the high audit probability it is optimal to use the self-disclosure option, whereas it is optimal not to use the option if the low audit probability is drawn. These options are highlighted with red lines. The last column presents the critical audit probabilities under which tax compliance becomes the optimal option. For the No Self-Disclosure Treatment under uncertainty subjects should evade all taxes along as the expected audit probability ($\frac{p_H+p_L^*}{\theta}$) does not exceed 50%.

Table 5.5: Share of Non-Evader per Period

Period	Audit probability	No Self-Disclosure (1)	Self-Disclosure without Penalty (2)	Self-Disclosure with Penalty (3)	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Total		38.33%	34.38%	40.00%	0.325	0.758	0.188
1-5		34.00%	37.50%	36.55%	0.609	0.715	0.899
6-12		41.43%	32.14%	41.87%	0.069	1.000	0.067
9&12	40% / 80%	81.67%	50.00%	63.79%	0.001	0.038	0.171
1	0%	6.67%	8.33%	0.00%	1.000	0.492	0.200
2	40%	16.67%	29.17%	31.03%	0.333	0.233	1.000
3	60%	56.67%	54.17%	62.07%	1.000	0.792	0.588
4	80%	86.67%	75.00%	79.31%	0.311	0.506	0.751
5	20%	3.33%	20.83%	10.34%	0.078	0.353	0.444
6	0% / 80%	23.33%	20.83%	34.48%	1.000	0.399	0.363
7	0% / 40%	6.67%	16.67%	10.34%	0.389	0.671	0.688
8	20% / 60%	36.67%	33.33%	44.83%	1.000	0.601	0.416
9	40% / 80%	86.67%	58.33%	65.52%	0.028	0.072	0.776
10	0% / 80%	33.33%	20.83%	41.38%	0.370	0.596	0.145
11	20% / 60%	26.66%	33.33%	34.48%	0.580	0.580	1.000
12	40% / 80%	76.67%	41.67%	62.07%	0.012	0.267	0.173

Notes: This table displays the share of NON-EVADER (i.e., participants who report all taxes in this period truthfully) per period and treatment. The first column presents the period(s), the second column the respective audit probabilities. The third (fourth; fifth) column displays the share of NON-EVADER in the No Self-Disclosure Treatment (Self-Disclosure without Penalty Treatment; Self-Disclosure with Penalty Treatment). Using 2-sided Fisher's exact tests we analyze whether the share of NON-EVADER significantly differ between the treatments. The respective p-values are presented in column 6 (7;8) for the comparison between the No Self-Disclosure Treatment and the Self-Disclosure without Penalty Treatment (No Self-Disclosure Treatment and Self-Disclosure with Penalty Treatment; Self-Disclosure without Penalty Treatment and Self-Disclosure with Penalty Treatment).

Table 5.6: Share of Full Evader per Period

Period	Audit probability	No Self-Disclosure (1)	Self-Disclosure without Penalty (2)	Self-Disclosure with Penalty (3)	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Total		22.22%	29.17%	22.99%	0.046	0.857	0.084
1-5		31.33%	29.17%	31.03%	0.790	1.000	0.789
6-12		15.71%	29.17%	17.24%	0.002	0.693	0.009
9&12	40% / 80%	1.67%	8.33%	6.90%	0.169	0.203	1.000
1	0%	86.67%	75.00%	89.66%	0.311	1.000	0.271
2	40%	16.67%	20.83%	20.69%	0.736	0.748	1.000
3	60%	0.00%	8.33%	6.90%	0.193	0.237	1.000
4	80%	0.00%	0.00%	0.00%			
5	20%	53.33%	41.67%	37.93%	0.425	0.299	1.000
6	0% / 80%	16.67%	33.33%	10.34%	0.206	0.706	0.050
7	0% / 40%	53.33%	41.67%	37.93%	0.425	0.299	1.000
8	20% / 60%	6.67%	29.17%	13.79%	0.062	0.424	0.194
9	40% / 80%	0.00%	4.17%	3.45%	0.444	0.492	1.000
10	0% / 80%	16.67%	54.17%	27.59%	0.008	0.360	0.089
11	20% / 60%	13.33%	29.17%	17.24%	0.186	0.731	0.341
12	40% / 80%	3.33%	12.5%	10.34%	0.312	0.353	1.000

Notes: This table displays the share of FULL EVADER (i.e., participants who evade taxes to the full extent) per period and treatment. The first column presents the period(s), the second column the respective audit probabilities. The third (fourth; fifth) column display the share of FULL EVADER in the No Self-Disclosure Treatment (Self-Disclosure without Penalty Treatment; Self-Disclosure with Penalty Treatment). Using 2-sided Fisher's exact tests we analyze whether the share of FULL EVADER significantly differ between the treatments. The respective p-values are presented in column 6 (7;8) for the comparison between the No Self-Disclosure Treatment and the Self-Disclosure without Penalty Treatment (No Self-Disclosure Treatment and Self-Disclosure with Penalty Treatment; Self-Disclosure without Penalty Treatment and Self-Disclosure with Penalty Treatment).

Table 5.7: Logit-Regression Results for Non-Evader

NON-EVADER	(A1)	(A2)	(A3)	(A4)
SELF-DISCLOSURE	-0.1832	0.0203	0.3204	0.4810
WITHOUT PENALTY	(0.8322)	(0.8014)	(0.9958)	(0.9645)
SELF-DISCLOSURE	0.2604	0.4819	0.2443	0.4357
WITH PENALTY	(0.6460)	(0.5207)	(0.7435)	(0.6203)
EXPECTED DIFFERENCE	1.7034***	1.8021***		
	(0.6491)	(0.6751)		
SELF-DISCLOSURE WITHOUT	-1.9159**	-2.0636**		
PENALTY X EXPECTED DIFFERENCE	(0.7898)	(0.8050)		
SELF-DISCLOSURE WITH	-1.4741*	-1.5457*		
PENALTY X EXPECTED DIFFERENCE	(0.8048)	(0.8336)		
UNCERTAIN PERIODS			1.0338**	1.0019**
			(0.4688)	(0.4685)
SELF-DISCLOSURE WITHOUT			-1.4969**	-1.4334**
PENALTY X UNCERTAIN PERIODS			(0.6950)	(0.6876)
SELF-DISCLOSURE WITH			-0.4121	-0.3953
PENALTY X UNCERTAIN PERIODS			(0.5688)	(0.5714)
ROUND	0.0062	0.0053	0.0141	0.0162
	(0.0606)	(0.0602)	(0.0567)	(0.0564)
AUDIT PROBABILITY	10.3995***	10.3579***	11.6539***	11.5753***
	(1.0386)	(1.0338)	(1.1547)	(1.1424)
AGE		0.0268		0.0251
		(0.0758)		(0.0803)
FEMALE		1.1870*		1.2118*
		(0.6547)		(0.6935)
ECONOMICS AND MANAGEMENT		-0.4118		-0.4481
		(0.5914)		(0.6335)
RISK AVERSION		0.2597**		0.2665*
		(0.1315)		(0.1393)
TAX EXPERIENCE		-0.3883*		-0.3963
		(0.2335)		(0.2513)
TAX MORALE		-0.1390		-0.1361
		(0.1263)		(0.1330)
POSITIVE RECIPROCITY		-0.0084		-0.0104
		(0.1419)		(0.1501)
NEGATIVE RECIPROCITY		-0.0189		-0.0228
		(0.1059)		(0.1123)
FAIRNESS		0.0157		0.0177
		(0.0975)		(0.1039)
COMPLEX DECISION		-0.6031***		-0.6055***
		(0.1392)		(0.1426)
CONSTANT	-5.6769***	-5.5505**	-6.6347***	-6.4404**
	(0.7079)	(2.6018)	(0.8651)	(2.7901)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.6074	0.5656	0.9396	0.9616
SDWP X EXPDIF = SDP X EXPDIF	0.4827	0.4026		
SDWP X UNC = SDP X UNC			0.0745	0.0849

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. The dependent dichotomous variable is NON-EVADER which takes the value 1 if the subject does not evade any taxes in the respective period and 0 otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE

WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. ROUND (AUDIT PROBABILITY) presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.8: Logit-Regression Results for Full Evader

FULLEVADER	(A5)	(A6)	(A7)	(A8)
SELF-DISCLOSURE	0.6555	0.1945	-0.2860	-0.7403
WITHOUT PENALTY	(0.6429)	(0.6034)	(0.7687)	(0.7187)
SELF-DISCLOSURE	0.0009	0.1050	-0.0158	0.0730
WITH PENALTY	(0.5820)	(0.6245)	(0.6669)	(0.7100)
EXPECTED DIFFERENCE	-0.7357	-0.6980		
	(1.1215)	(1.1152)		
SELF-DISCLOSURE WITHOUT	0.8028	0.7048		
PENALTY X EXPECTED DIFFERENCE	(1.2471)	(1.2428)		
SELF-DISCLOSURE WITH	1.1854	1.2022		
PENALTY X EXPECTED DIFFERENCE	(1.1866)	(1.1768)		
UNCERTAIN PERIODS			-0.4259	-0.4238
			(0.4046)	(0.3953)
SELF-DISCLOSURE WITHOUT			1.6459**	1.6242**
PENALTY X UNCERTAIN PERIODS			(0.7004)	(0.7028)
SELF-DISCLOSURE WITH			0.1124	0.1394
PENALTY X UNCERTAIN PERIODS			(0.5759)	(0.5692)
ROUND	0.0678	0.0671	0.0667	0.0652
	(0.0600)	(0.0602)	(0.0618)	(0.0620)
AUDIT PROBABILITY	-11.7200***	-11.7599***	-12.0622***	-12.0824***
	(1.1909)	(1.1911)	(1.1660)	(1.1685)
AGE		-0.0113		-0.0148
		(0.0751)		(0.0771)
FEMALE		-0.2276		-0.2449
		(0.5903)		(0.6141)
ECONOMICS AND MANAGEMENT		0.8908		0.8878
		(0.5668)		(0.5751)
RISK AVERSION		-0.1666		-0.1636
		(0.1080)		(0.1119)
TAX EXPERIENCE		0.0853		0.0957
		(0.2512)		(0.2568)
TAX MORALE		-0.0975		-0.1019
		(0.1020)		(0.1035)
POSITIVE RECIPROCITY		-0.1471		-0.1463
		(0.1410)		(0.1449)
NEGATIVE RECIPROCITY		0.0037		0.0076
		(0.1023)		(0.1053)
FAIRNESS		-0.0535		-0.0534
		(0.0845)		(0.0851)
COMPLEX DECISION		-0.1494		-0.1495
		(0.1022)		(0.1049)
CONSTANT	1.9710***	5.3545*	2.2691***	5.7240**
	(0.4965)	(2.7831)	(0.5176)	(2.8560)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.3288	0.8932	0.7348	0.2829
SDWP X EXPDIF = SDP X EXPDIF	0.6434	0.5414		
SDWP X UNC = SDP X UNC			0.0281	0.0356

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. The dependent dichotomous variable is FULL EVADER which takes the value 1 if the subject evades all taxes in the respective period and 0 otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. ROUND (AUDIT PROBABILITY)

presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.9: Regression Results for Tax Evasion without ROUND

TAX EVASION	(A9)	(A10)	(A11)	(A12)
SELF-DISCLOSURE	0.0697	0.0595	-0.0119	-0.0221
WITHOUT PENALTY	(0.0633)	(0.0745)	(0.0626)	(0.0735)
SELF-DISCLOSURE	0.0094	0.0213	-0.0118	0.0001
WITH PENALTY	(0.0541)	(0.0475)	(0.0492)	(0.0442)
EXPECTED DIFFERENCE	-0.1493***	-0.1493***		
	(0.0378)	(0.0380)		
SELF-DISCLOSURE WITHOUT	0.1502**	0.1502**		
PENALTY X EXPECTED DIFFERENCE	(0.0618)	(0.0621)		
SELF-DISCLOSURE WITH	0.0901*	0.0901*		
PENALTY X EXPECTED DIFFERENCE	(0.0501)	(0.0503)		
UNCERTAIN PERIODS			-0.1364***	-0.1364***
			(0.0313)	(0.0314)
SELF-DISCLOSURE WITHOUT			0.1827***	0.1827***
PENALTY X UNCERTAIN PERIODS			(0.0522)	(0.0524)
SELF-DISCLOSURE WITH			0.0621	0.0621
PENALTY X UNCERTAIN PERIODS			(0.0428)	(0.0430)
AUDIT PROBABILITY	-1.1003***	-1.1003***	-1.1431***	-1.1431***
	(0.0481)	(0.0483)	(0.0499)	(0.0502)
AGE		-0.0058		-0.0058
		(0.0066)		(0.0066)
FEMALE		-0.0638		-0.0638
		(0.0544)		(0.0544)
ECONOMICS AND MANAGEMENT		0.0075		0.0075
		(0.0519)		(0.0519)
RISK AVERSION		-0.0235**		-0.0235**
		(0.0109)		(0.0109)
TAX EXPERIENCE		0.0222		0.0222
		(0.0220)		(0.0220)
TAX MORALE		0.0026		0.0026
		(0.0099)		(0.0099)
POSITIVE RECIPROCITY		0.0082		0.0082
		(0.0116)		(0.0116)
NEGATIVE RECIPROCITY		0.0046		0.0046
		(0.0099)		(0.0099)
FAIRNESS		-0.0043		-0.0043
		(0.0086)		(0.0086)
COMPLEX DECISION		0.0179*		0.0179*
		(0.0093)		(0.0093)
CONSTANT	0.8677***	1.0208***	0.9403***	1.0934***
	(0.0387)	(0.1968)	(0.0417)	(0.1959)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.3923	0.6066	0.9997	0.7543
SDWP X EXPDIF = SDP X EXPDIF	0.3112	0.3136		
SDWP X UNC = SDP X UNC			0.0182	0.0188

Notes: The table presents the results of random-effects panel regressions which are clustered on subject level. The metric variable TAX EVASION serves as dependent variable and presents the ratio of tax payments that were initially not declared to the demanded tax payments. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. AUDIT PROBABILITY presents audit probability in percent in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the

Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.10: Logit-Regression Results for Non-Evader without ROUND

NON-EVADER	(A13)	(A14)	(A15)	(A16)
SELF-DISCLOSURE	-0.1833	0.0202	0.3240	0.4849
WITHOUT PENALTY	(0.8320)	(0.8013)	(0.9993)	(0.9670)
SELF-DISCLOSURE	0.2606	0.4820	0.2483	0.4401
WITH PENALTY	(0.6460)	(0.5208)	(0.7462)	(0.6235)
EXPECTED DIFFERENCE	1.7156***	1.8126***		
	(0.6407)	(0.6672)		
SELF-DISCLOSURE WITHOUT	-1.9158**	-2.0635**		
PENALTY X EXPECTED DIFFERENCE	(0.7897)	(0.8049)		
SELF-DISCLOSURE WITH	-1.4741*	-1.5457*		
PENALTY X EXPECTED DIFFERENCE	(0.8048)	(0.8336)		
UNCERTAIN PERIODS			1.0561**	1.0271**
			(0.4683)	(0.4688)
SELF-DISCLOSURE WITHOUT			-1.5009**	-1.4373**
PENALTY X UNCERTAIN PERIODS			(0.6983)	(0.6909)
SELF-DISCLOSURE WITH			-0.4168	-0.4005
PENALTY X UNCERTAIN PERIODS			(0.5703)	(0.5732)
AUDIT PROBABILITY	10.4061***	10.3637***	11.7148***	11.6448***
	(1.0416)	(1.0367)	(1.1731)	(1.1622)
AGE		0.0268		0.0250
		(0.0757)		(0.0804)
FEMALE		1.1868*		1.2123*
		(0.6545)		(0.6941)
ECONOMICS AND MANAGEMENT		-0.4116		-0.4490
		(0.5912)		(0.6340)
RISK AVERSION		0.2596**		0.2667*
		(0.1315)		(0.1394)
TAX EXPERIENCE		-0.3882*		-0.3964
		(0.2334)		(0.2515)
TAX MORALE		-0.1390		-0.1360
		(0.1263)		(0.1331)
POSITIVE RECIPROCITY		-0.0084		-0.0104
		(0.1418)		(0.1502)
NEGATIVE RECIPROCITY		-0.0189		-0.0228
		(0.1059)		(0.1124)
FAIRNESS		0.0158		0.0177
		(0.0975)		(0.1040)
COMPLEX DECISION		-0.6031***		-0.6057***
		(0.1391)		(0.1428)
CONSTANT	-5.6589***	-5.5349**	-6.6248***	-6.4276**
	(0.6768)	(2.5983)	(0.8603)	(2.7900)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.6071	0.5655	0.9401	0.9621
SDWP X EXPDIF = SDP X EXPDIF	0.4828	0.4026		
SDWP X UNC = SDP X UNC			0.0753	0.0861

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. The dependent dichotomous variable is NON-EVADER which takes the value 1 if the subject does not evade any taxes in the respective period and 0 otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. AUDIT PROBABILITY audit probability in percent in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and

Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.11: Logit-Regression Results for Full Evader without ROUND

FULL EVADER	(A17)	(A18)	(A19)	(A20)
SELF-DISCLOSURE	0.6529	0.1930	-0.2843	-0.7389
WITHOUT PENALTY	(0.6423)	(0.6030)	(0.7603)	(0.7108)
SELF-DISCLOSURE	-0.0008	0.1031	-0.0139	0.0753
WITH PENALTY	(0.5817)	(0.6243)	(0.6579)	(0.7020)
EXPECTED DIFFERENCE	-0.6411	-0.6044		
	(1.1015)	(1.0952)		
SELF-DISCLOSURE WITHOUT	0.8103	0.7133		
PENALTY X EXPECTED DIFFERENCE	(1.2455)	(1.2416)		
SELF-DISCLOSURE WITH	1.1889	1.2051		
PENALTY X EXPECTED DIFFERENCE	(1.1847)	(1.1750)		
UNCERTAIN PERIODS			-0.4294	-0.4273
			(0.3990)	(0.3905)
SELF-DISCLOSURE WITHOUT			1.6457**	1.6258**
PENALTY X UNCERTAIN PERIODS			(0.6943)	(0.6973)
SELF-DISCLOSURE WITH			0.1114	0.1373
PENALTY X UNCERTAIN PERIODS			(0.5695)	(0.5627)
AUDIT PROBABILITY	-11.4787***	-11.5235***	-11.7412***	-11.7715***
	(1.1507)	(1.1517)	(1.0511)	(1.0555)
AGE		-0.0120		-0.0151
		(0.0751)		(0.0770)
FEMALE		-0.2306		-0.2481
		(0.5901)		(0.6131)
ECONOMICS AND MANAGEMENT		0.8919		0.8886
		(0.5662)		(0.5740)
RISK AVERSION		-0.1666		-0.1636
		(0.1079)		(0.1117)
TAX EXPERIENCE		0.0852		0.0954
		(0.2504)		(0.2560)
TAX MORALE		-0.0975		-0.1015
		(0.1019)		(0.1033)
POSITIVE RECIPROCITY		-0.1471		-0.1467
		(0.1412)		(0.1448)
NEGATIVE RECIPROCITY		0.0039		0.0075
		(0.1022)		(0.1050)
FAIRNESS		-0.0527		-0.0530
		(0.0843)		(0.0850)
COMPLEX DECISION		-0.1486		-0.1490
		(0.1024)		(0.1049)
CONSTANT	2.1142***	5.5076**	2.3916***	5.8506**
	(0.4890)	(2.7767)	(0.5214)	(2.8484)
Observations	996	996	996	996
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.3288	0.8927	0.7316	0.2776
SDWP X EXPDIF = SDP X EXPDIF	0.6458	0.5447		
SDWP X UNC = SDP X UNC			0.0271	0.0340

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. The dependent dichotomous variable is FULL which takes the value 1 if the subject evades all taxes in the respective period and 0 otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. AUDIT PROBABILITY audit probability in percent in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise.

RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.12: Regression Results for Tax Evasion with LAST PERIOD AUDIT

TAX EVASION	(A21)	(A22)	(A23)	(A24)
SELF-DISCLOSURE	0.0871	0.0770	0.0036	-0.0065
WITHOUT PENALTY	(0.0654)	(0.0788)	(0.0676)	(0.0807)
SELF-DISCLOSURE	0.0080	0.0226	-0.0239	-0.0092
WITH PENALTY	(0.0586)	(0.0514)	(0.0577)	(0.0516)
EXPECTED DIFFERENCE	-0.1517***	-0.1518***		
	(0.0397)	(0.0399)		
SELF-DISCLOSURE WITHOUT	0.1296**	0.1297**		
PENALTY X EXPECTED DIFFERENCE	(0.0622)	(0.0625)		
SELF-DISCLOSURE WITH	0.0889*	0.0890*		
PENALTY X EXPECTED DIFFERENCE	(0.0518)	(0.0521)		
UNCERTAIN PERIODS			-0.1375***	-0.1376***
			(0.0358)	(0.0360)
SELF-DISCLOSURE WITHOUT			0.1683***	0.1684***
PENALTY X UNCERTAIN PERIODS			(0.0574)	(0.0577)
SELF-DISCLOSURE WITH			0.0756*	0.0756*
PENALTY X UNCERTAIN PERIODS			(0.0448)	(0.0451)
ROUND	0.0025	0.0025	-0.0014	-0.0014
	(0.0055)	(0.0056)	(0.0051)	(0.0052)
AUDIT PROBABILITY	-1.0554***	-1.0554***	-1.1556***	-1.1557***
	(0.0712)	(0.0716)	(0.0714)	(0.0718)
LAST PERIOD AUDIT	0.0202	0.0199	0.0208	0.0206
	(0.0212)	(0.0211)	(0.0222)	(0.0221)
AGE		-0.0052		-0.0052
		(0.0069)		(0.0069)
FEMALE		-0.0616		-0.0616
		(0.0572)		(0.0572)
ECONOMICS AND MANAGEMENT		0.0006		0.0006
		(0.0543)		(0.0544)
RISK AVERSION		-0.0245**		-0.0245**
		(0.0113)		(0.0113)
TAX EXPERIENCE		0.0249		0.0249
		(0.0233)		(0.0233)
TAX MORALE		0.0040		0.0040
		(0.0105)		(0.0105)
POSITIVE RECIPROCITY		0.0121		0.0121
		(0.0121)		(0.0121)
NEGATIVE RECIPROCITY		0.0060		0.0060
		(0.0107)		(0.0107)
FAIRNESS		-0.0059		-0.0059
		(0.0091)		(0.0091)
COMPLEX DECISION		0.0192**		0.0192**
		(0.0098)		(0.0098)
CONSTANT	0.8251***	0.9265***	0.9451***	1.0465***
	(0.0519)	(0.2084)	(0.0594)	(0.2092)
Observations	913	913	913	913
Number of Subject	83	83	83	83
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Wald Test (Prob > chi2):				
SDWP = SDP	0.2920	0.4919	0.7073	0.9726
SDWP X EXPDIF = SDP X EXPDIF	0.5013	0.5036		
SDWP X UNC = SDP X UNC			0.0874	0.0892

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. The dependent dichotomous variable is NON-EVADER which takes the value 1 if the subject does not evade any taxes in the respective period and 0 otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY measure treatment effects and take the value 1 (0) if the subject is (not) granted the self-disclosure without (with) penalty opportunity. EXPECTED DIFFERENCE takes the value 1 for periods 9 and 12 where we expect to find a divergent tax evasion behavior and zero otherwise. SELF-DISCLOSURE WITHOUT (WITH) PENALTY X EXPECTED DIFFERENCE are the interaction terms of both dummy variables and present the influence of the respective treatment on tax evasion in periods 9 and 12. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-

DISCLOSURE WITHOUT (WITH) PENALTY X UNCERTAIN PERIODS are the interaction terms of both dummy variables and present the influence of the treatment effect on tax evasion in uncertain periods. ROUND (AUDIT PROBABILITY) presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. LAST PERIOD AUDIT is a dummy variable which takes the value 1 if the subject was audited in the previous period and 0 otherwise. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. The last three columns present results for different Wald tests run after each regression. SDWP = SDP compares the coefficients of the Self-Disclosure without Penalty Treatment and the Self-Disclosure with Penalty Treatment. SDWP X EXPDIF = SDP X EXPDIF (SDWP X UNC = SDP X UNC) compares the coefficients of the respective interaction terms SELF-DISCLOSURE WITHOUT PENALTY X EXPECTED DIFFERENCE and SELF-DISCLOSURE WITH PENALTY X EXPECTED DIFFERENCE (SELF-DISCLOSURE WITHOUT PENALTY X UNCERTAIN PERIODS and SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5.13: Logit-Regression Results for the Self-Disclosure Use after Initial Tax Evasion

SELF-DISCLOSURE USE	(A25)	(A26)	(A27)	(A28)
SELF-DISCLOSURE WITH PENALTY	1.4931**	1.8024***	0.2191	0.0913
	(0.7046)	(0.6502)	(1.2471)	(1.3170)
UNCERTAIN PERIODS	1.7235***	1.6917***	1.5245*	1.3363**
	(0.5002)	(0.4983)	(0.7790)	(0.5994)
SELF-DISCLOSURE WITH PENALTY	-1.2427*	-1.2215*	-0.6542	-0.3828
X UNCERTAIN PERIODS	(0.7218)	(0.7064)	(1.3612)	(1.2184)
ROUND	-0.2443***	-0.2469***	-0.1529	-0.1882
	(0.0764)	(0.0773)	(0.1463)	(0.1473)
AUDIT PROBABILITY	5.0648***	5.1451***	6.4127***	7.2500***
	(1.0004)	(0.9872)	(2.2253)	(2.2417)
AGE		0.1111		0.0491
		(0.0829)		(0.1623)
FEMALE		0.6393		0.9525
		(0.6134)		(1.0022)
ECONOMICS AND MANAGEMENT		-0.2778		-0.2892
		(0.3843)		(0.6706)
RISK AVERSION		-0.1221		-0.1806
		(0.1059)		(0.1658)
TAX EXPERIENCE		0.0741		0.4873*
		(0.1926)		(0.2725)
TAX MORALE		0.0050		0.3550**
		(0.0861)		(0.1525)
POSITIVE RECIPROCITY		-0.0084		-0.2149
		(0.1046)		(0.3174)
NEGATIVE RECIPROCITY		0.0270		-0.1659
		(0.0674)		(0.1477)
FAIRNESS		-0.0176		0.1553
		(0.0895)		(0.1389)
COMPLEX DECISION		-0.1006		-0.1182
		(0.1179)		(0.1835)
CONSTANT	-4.0819***	-6.1940***	-4.2801***	-6.0810
	(0.7384)	(2.1516)	(1.1751)	(5.6353)
Observations	399	399	164	164
Number of Subject	51	51	49	49
Prob > chi2	0.0000	0.0007	0.0237	0.3878

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. While models A25 and A26 analyze all relevant observations, models A27 and A28 only test for full evading decisions. The dichotomous variable SELF-DISCLOSURE USE serves as dependent variable and takes the value 1 if the subject uses the self-disclosure option after an initial tax evasion and zero otherwise. UNCERTAIN PERIODS takes the value 1 if the subject is confronted with two possible audit probabilities (i.e., in the last seven periods) and zero otherwise. SELF-DISCLOSURE WITH PENALTY X UNCERTAIN PERIODS is the interaction term of both dummy variables and presents the influence of the treatment effect in uncertain periods. ROUND (AUDIT PROBABILITY) presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5.14: Logit-Regression Results for the Self-Disclosure Use after Initial Tax Evasion Testing for the Occurrence of High Audit Probabilities

SELF-DISCLOSURE USE	(A29)	(A30)	(A31)	(A32)
SELF-DISCLOSURE WITH PENALTY	-0.1828	0.4896	-6.1102	-4.8062
	(0.6407)	(0.6703)	(10.1699)	(14.9423)
ROUND	-0.1965*	-0.1925*	0.2027	0.1877
	(0.1112)	(0.1090)	(0.5479)	(0.6851)
HIGH AUDIT PROBABILITY	2.9693***	2.9275***	10.5995**	8.2965
	(0.7712)	(0.7934)	(5.3324)	(15.9866)
AUDIT PROBABILITY	3.0945	3.1439	-19.4136	-15.3495
	(2.4321)	(2.4464)	(19.9125)	(35.5056)
AGE		0.0216		-0.1234
		(0.1174)		(1.1794)
FEMALE		0.0836		4.7485
		(0.8521)		(14.6056)
ECONOMICS AND MANAGEMENT		0.3101		4.8313
		(0.5612)		(11.7495)
RISK AVERSION		0.0216		0.9065
		(0.1421)		(2.4089)
TAX EXPERIENCE		0.3077		2.6567
		(0.2584)		(6.6374)
TAX MORALE		-0.1505		0.4117
		(0.1269)		(1.4125)
POSITIVE RECIPROCITY		0.0915		0.9797
		(0.1872)		(3.4523)
NEGATIVE RECIPROCITY		-0.0105		-0.0516
		(0.1175)		(1.3353)
FAIRNESS		0.0659		0.6038
		(0.1220)		(1.8181)
COMPLEX DECISION		-0.2256		-1.6110
		(0.1775)		(3.6729)
CONSTANT	-3.4527***	-4.4681	0.2470	-21.0493
	(1.2581)	(3.6886)	(7.1069)	(51.0869)
Observations	186	186	63	63
Number of Subject	44	44	28	28
Prob > chi2	0.0000	0.0099	0.2075	0.9997

Notes: The table presents the results of random-effects logit regressions which are clustered on subject level. While models A29 and A30 analyze all relevant observations, models A31 and A32 only test for full evading decisions. The dichotomous variable SELF-DISCLOSURE USE serves as dependent variable and takes the value 1 if the subject uses the self-disclosure option after an initial tax evasion and zero otherwise. HIGH AUDIT PROBABILITY takes the value 1 (0) if the high (low) audit probability is drawn. ROUND (AUDIT PROBABILITY) presents the round taking account of the introduction of two audit probabilities by starting over counting at 1 in period 6 (audit probability in percent) in which the tax evasion decision is made. AGE is measured in years. FEMALE (ECONOMICS AND MANAGEMENT) is a dummy variable and takes the value 1 if the subject is female (studies at the Faculty of Economics and Management) and 0 otherwise. RISK AVERSION (TAX MORALE) measures the subject's self-assessed risk aversion (tax morale) on an 11-point scale (a 10-point scale). TAX EXPERIENCE is measured on a 4-point scale and presents the subject's experience with real tax filings. POSTIVE (NEGATIVE) RECIPROCITY is measured on an 11-point scale. COMPLEX DECISION (FAIRNESS) reflects the subject's opinion on the decision periods' complexity (on the fairness of the contribution's and control's systems) and is measured on an 11-point scale. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 6

Tax Rate Complexity and Labor Supply – A Real-Effort Experiment[*]

Abstract

This paper analyses the influence of tax rate complexity on individual labor supply. In a real-effort experiment, subjects decide between a real work task to earn real money or a leisure alternative. In four different tax rate complexity treatments, we measure the individual labor supply of each subject. The results show that taxpayers' labor supply decreases with increasing tax rate complexity.

Keywords: Real-Effort Experiment Tax Rate Complexity · Work-Leisure- Decision

JEL Codes: C91 · H24 · H31

[*] This chapter is co-authored by Dr. Christian Sielaff (Freie University Berlin) and was published in Die Betriebswirtschaft/ Business Administration Review (2016), 65-84.
We are grateful to Kay Blaufus, Frank Hechtner, Jochen Hundsdoerfer and to two anonymous referees for their helpful comments and advice. Furthermore, we are also grateful for the support of the student assistants from the department of Accounting, Auditing and Taxation of the Freie Universität Berlin when conducting the experiment.

6.1 Introduction

The complexity of the German income tax system is of interest in numerous political and academic discussions, particularly the question of fiscal simplification which is constantly demanded by different political, economic and academic stakeholders. However, concrete behavioral responses of taxpayers that are induced by a complex tax system are often neglected. Thus, this paper aims to examine this issue. Through the use of an experiment, we analyze the influence of a tax system's complexity on taxpayers' labor supply decisions.

In the experiment, the participants are confronted with a real effort-task in which they are asked to transfer answer sheets from a multiple choice test to an Excel spreadsheet. They are free to stop this real work task at any time and to choose a real leisure alternative. Each participant receives the same gross wage per correctly transferred answer sheet, which is subject to taxation. The participants are assigned to four different tax systems (between-subject design) that differ in terms of tariff complexity. However, the effective tax rate and hence the net wage are identical in all of the treatments. The tax system's complexity is modelled by the interaction of different tax rates. Thus, we intend to analyze the research question of whether and how tax system's complexity influences a taxpayer's tax perception and consequently her individual labor supply. We chose to model tax complexity through the interaction of different tax rates because it reflects numerous cases in the national[1] and international[2] tax tariff system and affects nearly all taxpayers.

We expect participants to be increasingly incapable of calculating their effective tax rates with growing tax rate complexity. Rather, they will use heuristics or simplification rules to estimate their tax burden. This paper focuses on analyzing whether these heuristics lead to an over- or underestimation of the actual tax burden, as the directional effects are crucial for either the extension or reduction of the labor supply. We show that taxpayers react to increasing tax complexity with a reduction of the labor supply. Hence, tax tariff complexity leads to a systematic deviation from the theoretical prediction of an accurately calculating taxpayer. This reaction to tax tariff complexity is essential to effective discussions about the necessity of complexity reduction within the tariff structure. An experiment is an appropriate setting to examine this issue because it allows for the analysis of taxpayers' behavior under controlled conditions. Monetary incentives create a convincing real-effort decision under the influence of real tax

[1] In the national context, the combination of income tax, solidarity surcharge and church tax as well as the interaction of income tax, trade tax and the trade tax' deduction according to § 35 EStG are examples of tax rate complexity through the interaction of different tax rates. This tax rate complexity is often discussed in political debates besides simplifying the tax base. In this context, the self-deductibility of the trade tax was abolished in the 2008 corporate tax reform, which simplified tax calculations. Furthermore, the FDP (liberal party) proposed the introduction of a bracket model that reduces tax complexity through tax rate simplifications (FDP, 2010, p. 6.).

[2] In the international context, the combination of local taxes and federal taxes in the U.S. as well as the interaction of domestic and foreign taxes are classic examples of tax rate complexity.

payments. Hence, the experimental results can be used to derive predictions on the impact of tax simplifications.

This study is relevant for both theoretical economic research, especially the field of behavioral taxation, as well as for continuous political discussion. Prior theoretical research generally assumes that taxpayers fully understand their tax tariff systems and are thus able to rationally utilize tax provisions for tax planning purposes. Thus, Ramsey (1927) suggests that taxpayers respond to tax changes similarly to price changes. These models are therefore used to predict actual behavior. However, if the majority of taxpayers are not able to perceive tax issues appropriately due to tax tariff complexity, the model's explanatory and forecasting power diminishes. This study examines the issue by experimentally analyzing whether taxpayers demonstrate any behavioral constraints if the tax tariff system is not fully understood due to high complexity. Accordingly, behavioral reactions should be included in the economic models.

In the context of fiscal discussions, politicians often address tax simplification. However, the tax tariff system has become increasingly complicated due to special tax law provisions. Currently, in addition to the basic tariff regulations in § 32a EStG, there are five additional income tariff standards that increase the income tax system's complexity (Hechtner, 2010, p. 27). Policymakers should consider taxpayers' responses to tax complexity and tax simplification when deciding on future tax regulations if tax tariff complexity, in addition the level of taxation, is essentially influencing taxpayers' behavior and decisions.

The paper is organized as follows: In Section 6.2, we provide a literature review on tax complexity and the influence of taxes on the labor supply. In Section 6.3, we first derive our hypothesis and specify our research question, and second, we describe our experiment. In Section 6.4, we present our experimental results. Our results are interpreted and discussed in Section 6.5. Section 6.6 presents the study's limitations and avenues for further research.

6.2 Literature Review

Our study links with the previous literature on the perception and impact of complex tax tariff systems and expands on this issue with a new research question and a realistic experimental model of both labor-leisure decisions and tax tariff complexity. For classification purposes, we present a literature overview that first focuses on tax complexity and second presents findings on the impact of taxes on the labor supply.

The literature on tax complexity and its impacts is broad and manifold. Congdon et al. (2009) find that taxpayers do not respond to actual but to individually perceived tax rates. Because tax tariff complexity increases the range of tax rate interpretations, studies should attach more value to the perceived tax rate to examine tax impacts. The present paper takes up this idea and investigates the influence of perceived tax rate complexity and its effect on economic decisions. Blaufus and Ortlieb (2009) study the impact of tax complexity by conducting a conjoint analysis. They measure tax complexity by the time needed

to understand a tax tariff system. These systems vary in lengths of descriptions and the number of technical (tax) terms. Blaufus and Ortlieb (2009) show that increasing complexity decreases the number of taxpayers who consider taxes in their decisions. Applying these findings to our study, the failure to calculate tax payments due to rising complexity would lead to an extension of the labor supply if taxpayers use the fixed gross wage as a heuristic.

Besides these studies, there are other current papers that investigate the impact of tax complexity on investment behavior. Ackermann et al. (2013), Ackermann et al. (2016) as well as Fochmann and Hemmerich (2014) find a perception effect that leads to decreasing risky investment after a tax introduction, even if the net payments are equal. The authors argue that the introduction of taxation increases the decision's complexity, which in turn decreases risk-taking. Hence, in accordance with our research question, these papers also show that taxes increase complexity, which leads to adaptation reactions.

Regarding the presentation of taxes, Chetty et al. (2009) conduct a field study and find that taxes are only perceived correctly by taxpayers if they are salient. They find that posting tax-inclusive price tags in a grocery store decreases the demand in comparison with stores where the sales tax is added to the bill only at the register. Finkelstein (2009) obtains similar results with an examination of tolls. She proves that drivers' price elasticity is reduced after the introduction of electronic toll collection, as this payment method is less salient. These observations also link to our study, as complexity induced by the interaction of tax rates may also cause a salience effect. On the one hand, taxpayers' awareness of taxes might be raised due to a longer and more complex description of the tax tariff system; on the other hand, they might be less capable of correctly estimating their tax payments.

In addition to these studies on tax complexity, there is another literature strand that focusses on determinants of labor supply. In accordance with Chetty et al. (2009), there are further studies that examine an influence of gross wage illusion on decision making. Keeping the net wage constant, Djanali and Sheehan-Connor (2012) find that labor supply is greater if a tax is levied on a higher gross wage than in the absence of taxation. Accordingly, under constant net wages Fochmann et al. (2013) reveal that labor supply is higher if participants face a higher gross wage, yet also a higher tax rate. This misperception of the net wage is explained with net wage illusion. While the model of the work-leisure decision is comparable to our experiment, the tax system model is not. Whereas Fochmann et al. (2013) vary the gross wage and the tax rate to examine behavioral effects, we only model tax tariff complexity through the interaction of different tax rates. Therefore, we can eliminate net wage illusion, tax base effects and the impact of gross wage, tax rates and net wage differences.

In the context of gross wage illusion, Hayashi et al. (2012) study the influence of net wage presentation on labor supply. They find that the labor supply is higher if participants are paid a higher gross wage and taxes are levied on it than if they are paid a lower wage and receive additional bonus payments. While Hayashi et al. (2012) model the leisure alternative by offering entertainment videos to watch,

other studies offer computer games, music or consuming specific goods. However, in our experiment, we offer a real leisure alternative to create a realistic environment and thus increase the study's validity (in accordance with Charness and Kuhn, 2011, pp. 239f.). Hence, we offer participants the opportunity to end the experiment at any time to pursue actual leisure activities.

There are further studies that discuss the impact of tax rates on the labor supply. Blumkin et al. (2012) conduct a real-effort experiment in which participants must invest their earned income into offered goods to examine the influence of an income tax and a consumption tax on labor supply. They find that an income tax reduces the labor supply more than an equivalent consumption tax. Weber and Schram (2017) conduct a real-effort experiment in which equivalent taxes are either paid directly by the employees (income tax) or by the employers (labor market taxes). They find that the labor supply is lower if taxes are paid by employers as labor market taxes than if they are paid directly by the employees as income taxes. Hence, participants differently evaluate tax systems due to presentation differences, although the tax rates and payments are equal.

Sillamaa (1999) conducts a replication study of Swenson (1988) to examine the influence of different linear tax rates on the labor supply. She confirms theoretical predictions that increasing tax rates and thus lower net wages decrease the labor supply. Transferring these results to our research question, we would expect participants to lower their labor supply if tax tariff complexity leads to an overestimation of tax rates. In a field study, Chetty and Saez (2013) investigate whether providing simple, personalized information on the Earned Income Tax Credit changes the labor supply in subsequent years. They find that no form of information, i.e., verbal, graphical, or tabular descriptions, leads to a respective labor supply response. As taxpayers do not accurately respond to tax credits, our experiment only focusses on tax payments and tax tariff complexity.

Dickinson (1999) experimentally examines labor supply responses to wage changes. While one group can only adjust their performance, a second group can additionally adjust their labor supply via working time. Thus, participants in the second group have the same opportunities to respond to wage changes as in our experiment: they can either work slower or faster or they can leave the experiment and choose a real leisure alternative. While Dickinson (1999) finds substitutional effects of labor performance for the first group, he finds both off- and on-the-job leisure responses for the second group. Thus, we expect our participants to respond similarly by either adjusting the labor supply or labor performance. In a recent study that focuses on a very similar research question to that in the present paper, Abeler and Jäger (2015) analyze the tax system's influence on the optimal labor supply. They design a real-effort experiment in which it is only worth working until a payoff-maximizing output level, at which time further labor supply yields negative net payments. The authors show that participants do not adjust their labor supply that they have chosen under a more simple tax system strongly enough. In contrast to our study, Abeler and Jäger (2015) model an optimal output from which further labor supply is no longer profitable. In our experiment, additional labor always yields additional positive payments. Furthermore,

we do not study responses to tax complexity changes but the impact of tax tariff complexity on labor supply and effort. Thus, both studies extend the research field on tax complexity and its impacts on the labor supply in different ways.

6.3 Hypothesis and Experimental Design

6.3.1 Research Question and Hypothesis

In an environment in which a rational taxpayer is always able to correctly anticipate tax payments, the tax system's design or complexity should not impact the accurate calculation of the effective tax rate. Thus, taxpayers should offer the same labor supply, independent of the tax tariff complexity. However, the presented literature has already demonstrated that taxpayers are not necessarily able to correctly determine their tax payments. In particular, those taxpayers who face a complex tax system make use of heuristics and other simplification rules to anticipate their tax burden. Hence, they either under- or overestimate their actual tax payments and their net wage. Therefore, taxpayers have to adjust their labor supply decisions based on these biased values.

We expect fewer taxpayers to be able to calculate their effective tax rate under increasing tax tariff complexity. Thus, they might use heuristics to estimate their effective tax rate, such as the anchor effect (Tversky and Kahneman, 1974), which drives taxpayers to orientate toward the highest or first mentioned tax rate (and disregard all other tax rates) or toward the sum of all tax rates. This expectation leads us to our first hypothesis:

Hypothesis 1: *Increasing tax system complexity decreases the number of taxpayers who are able to correctly determine their tax burden.*

It is of special interest whether the heuristics used by the taxpayers result in over- or underestimation of the actual tax burden and the net wage, as this direction significantly influences labor supply decisions. However, the direction of the effect of an over- or underestimation of the effective tax rate on the labor supply cannot be determined unambiguously beforehand because it can lead to either an income or substitution effect. If taxpayers overestimate the tax rate and thus underestimate the net wage, the income effect can either decrease (because labor is less valued) or increase the labor supply (if taxpayers want to earn a certain minimum wage). The substitution effect, however, always leads to a reduction of the labor supply as under decreasing wages, working time is substituted by leisure. It depends on the taxpayer's individual behavior whether the income and substitution effects act in the same direction or whether the income effect (over-)compensates the substitution effect. Prior studies reveal a positive relation between net wage and the labor supply. Thus, we accordingly expect that the (over-) underestimation of the tax burden and the (under-) overestimation of the respective net wage lead to an extension (reduction) of the labor supply (for theoretical predictions, see Hicks, 1935; for empirical studies on the impact of wage changes on labor supply, see Break, 1957; Hausman, 1981; Killingsworth,

1983; Kimmel and Kniesner, 1998; Ziliak and Kniesner, 1999; and Keane, 2011). Hence, hypothesis 2 addresses the consequences of the misperception of the tax rate on the labor supply:

Hypothesis 2: Increasing tax system complexity leads to an adapted labor supply.

The tax system's complexity can also impact working morality if payment is the work's major incentive. If the payment cannot or can only be imperfectly determined beforehand, work performance is affected. Thus, hypothesis 3 on work performance is as follows:

Hypothesis 3: Increasing tax system complexity leads to an adapted work performance.

This study focuses on shedding light on the direction and quantification of taxpayers' reactions to tax system complexity. Beforehand, we cannot theoretically predict whether the income effect outweighs the substitution effect or vice versa. Thus, we use an experiment to test our research question. Additionally, we provide indications for the used heuristics. Therefore, the tax tariff systems were modelled in such a way as to analyze these simplification rules.

6.3.2 Experimental Protocol

The experiment was conducted in five sessions in the computerized rooms of the School of Business and Economics of the Freie University Berlin. In total, 96 students participated in the experiment. The students' attention was drawn to the experiment by announcements before several lectures at the School of Business and Economics. Students registered for the experiment by providing their email address and session preferences. Students from all faculties applied for the experiment. Most of the participants were students in their second semester pursuing a bachelor's degree in economics.[3] However, because there are no significant differences between the sessions or between economics and non-economics students, we can aggregate all of the groups for further analysis. When announcing the experiment, the students were informed that the experiment would last at least three hours. Thus, we intended to prevent short-term appointments that could have caused an early termination of the experiment independent of the experiment's work task. Further information on the experiment, e.g., the research question, was not provided beforehand to prevent a self-selection bias.

The participants were randomly assigned to personal computers that were separated by a visual cover. Furthermore, communication was forbidden at any time during the experiment. A maximum of 25 participants attended each session and several experimenters supervised the sessions and answered any questions. At the beginning of the experiment, participants were provided with instructions that contained all of the necessary information about the experiment's procedure. Participants were provided

[3] Recent literature has rejected the argument that students do not serve as appropriate experimental participants; see Plott (1987) and Alm (2010). As our research question aims to analyze participants' general behavior, students appear to be a representative group; see Charness and Kuhn (2011), p. 234. Further studies also reveal that students are a reliable surrogate for the general population, particularly because they are confronted with similar underlying decision situations when they leave school; see Gemünden (1985) or Ashton and Kramer (1980).

enough time to read the instructions thoroughly and to ask questions, which were answered by the experimenters. Subsequently, the work task was exemplarily presented on the overhead projector and the experimenters explained how the answer sheets have to be transferred. By posing specific queries, we ensured that there were no further questions or clarifications concerning the experiment's procedure and the work task.

6.3.3 Experimental Design

Participants are confronted with a realistic work-leisure decision. Their task is to transfer answer sheets from a multiple choice test to an Excel spreadsheet for correction purposes. Every answer sheet consists of 40 questions with six answer possibilities, for which one is always checked. By choosing this work task, we intend to put the participants in a realistic working environment that requires neither special skills nor knowledge. The participants earn 0.30 Euro for every correctly transferred answer sheet. Their leisure alternative is to leave the experiment at any time and pursue actual leisure activities. Thus, ours is one of the few experiments that creates a realistic leisure opportunity. Hence, we generate a decision situation that is close to reality. Participants are told that there is no time limit, so they are free to work as long as they want.[4] In every session, there are two specially briefed participants whose task it is to leave the experiment after approximately thirty minutes to eliminate any possible inner obstacles to an early termination of the experiment. However, we do not observe any contagion effects, i.e.; actual participants leaving immediately after the two briefed participants.

The participants are informed that they have to pay taxes on their gross wage. After reading the instructions, each participant is randomly assigned to one of the four tax system sheets that explain the respective tax system, whereby the tax systems are equally distributed in every session.[5] Thus, we can control for session specific variations. Each of the four tax system sheets represents one treatment (between-subject design). However, participants were not informed about the net wage itself. The tax systems differ in level of complexity according to an increasing number of interacting tax rates. Table 6.1 provides an overview on the tax rates used in each tax system.

[4] The experiment was terminated after approximately three hours without previous announcement. Only five participants over all of the sessions were still working at this time. All of the other participants left the experiment without any knowledge about the termination. Some may argue that the participants were deceived. However, we choose to disregard this issue because it only affected five participants and because the participants were informed beforehand that the experiment would last "at least three hours".

[5] The instructions are presented in Section 6.8 (Appendix A), the tax system sheets are presented in Section 6.81 (Appendix B) and the questionnaire is displayed in Section 6.9 (Appendix C).

Table 6.1: Overview of Tax Tariff Systems

Tax System	Income Tax	Additive Tax on the Income	Surcharge Tax on the Income Tax	State Tax on the Income	Community Tax with Deductibility for Income Tax	Effective Tax Rate	Modelling of Complexity
A	60%	---	---	---	---	60.0%	simple tax rate
B	40%	20%	---	---	---	60.0%	additive linkage
C	40%	---	20%	12%	---	60.0%	additive and multiplicative linkage
D	20%	---	20%	20%	20%	59.2%	additive and multiplicative linkage and deductibility

Note: The treatments are modelled as between-subject design in which participants are randomly assigned to one of the four tax systems.

While in Tax System A, a flat tax of 60% (income tax) is levied, participants have to sum up two taxes (an income tax and an additive tax) in Tax System B to determine their effective tax rate (40% + 20% = 60%). In Tax System C, participants additionally have to apply a multiplicative linkage (40% · (1 + 20%) + 12% = 60%) to connect three different tax rates (income tax, surcharge tax and state tax). The most complex tax system, System D, further includes the deductibility of a fourth tax (community tax) for the tax base of the income tax ([20% · (1 + 20%)] · (1 – 20%) + 20% + 20% = 59.2% ≈ 60%). Thus, it is notable that we generate increasing complexity through an increasing number of tax rates and their interaction while keeping the effective tax rate at a constant rate of 60%.[6] We deliberately chose a rather high tax rate to sensitize the participants for tax payments.

When modelling the increasing tax system complexity, we use mathematical interactions that can also be found in German tax law. We keep the language on the tax system sheets as simple as possible and try to avoid legal and technical terms. However, when labelling the different tax rates on the tax system sheets, we also try to eliminate labelling or affectation effects through the use of real terms. With the intention of finding the right balance of terminology, we do not abstract the terminology too much (by calling it tax 1, tax 2 etc.) so we do not create an absolutely theoretical framework. Nonetheless, it is debatable whether the different labelling of taxes between the treatments may have caused confounders that cannot be measured. Yet, we try to label taxes similarly that work mathematically equally to

[6] In Tax System D, the effective tax rate is 59.2% because an effective tax rate of 60% in System D would have required non-integer single tax rates. We chose to refrain from non-integer values for simplification purposes and to prevent other undesired effects, such as a salience effect. However, if participants actually calculated the effective tax rate, the difference of 0.8%-points is rather low and negligible.

decrease this potential effect. We purposely waive modelling tax complexity by using a progressive tax system, as this additional trigger would be counterproductive and is already widely analyzed in the literature by examining that subjects use their average tax rate rather than the marginal tax rate when making economic decisions (see, e.g., De Bartolome, 1995; Gensemer et al., 1965; Lewis, 1978; Fujii and Hawley, 1988; Rupert and Fisher, 1995; Rupert and Wright, 1998; Rupert et al., 2003; Boylan and Frischmann, 2006; as well as Hundsdoerfer and Sichtmann, 2007). Furthermore, participants are not informed about the taxes' usage. Thereby, we want to eliminate affection effects (see Hundsdoerfer et al., 2013). The faculty actually benefitted indirectly from the taxes, as they decrease the experiment's costs. However, there has not been any query about this issue.

Table 6.2: Descriptive Statistic in Total and Separated by the Tax Systems

treatment group	A mean (SD)	B mean (SD)	C mean (SD)	D mean (SD)	Total mean (SD)
OBSERVATIONS	24	24	24	24	96
GENDER (FEMALE)	66.67%	58.33%	58.33%	50.00%	58.33%
AGE	23.00 (0.482)	23.46 (3.856)	24.63 (6.170)	23.08 (2.483)	23.54 (4.284)
ECONOMICS MAJOR	79.17% (0.415)	75.00% (0.442)	79.17% (0.415)	83.33% (0.381)	79.17% (0.408)
SEMESTER	3.91 (2.729)	3.82 (2.481)	3.64 (2.279)	3.96 (2.305)	3.83 (2.418)
TAX KNOWLEDGE	4.50 (1.719)	4.04 (2.386)	4.17 (2.513)	3.79 (2.146)	4.13 (2.192)
TAX RATE CALCULATION	87.50%	83.33%	58.33%	37.50%	66.67%

Note: In the table, the mean values and standard deviations (SD) of the descriptive variables, if relevant, are presented for each single tax systems and in total. The number of observations is an absolute number. GENDER (ECONOMICS MAJOR) denotes whether the participant is female (studies economics). Otherwise, the participant is male (takes another course of study). TAX KNOWLEDGE mirrors the participant's self-reported proficiency concerning taxes (metered on a 10-point scale where 1 = no knowledge and 10 = wide knowledge). TAX RATE CALCULATION denotes, whether the participant claims to have calculated her tax rate beforehand. There are no significant differences in the descriptive variables between the treatments.

After being assigned to their respective tax system, the participants receive a sufficient number of answer sheets and start the experiment simultaneously whereby the starting time is captured for evaluation by the experimenter. The answer sheets are designed in such a way that the participants are not able to realize any scheme that can ease the work task. No further tools are provided. After the participant ends the work task, the experimenter captures the finishing time and directs her to a separate room for payment. The participant's working time and performance (measured by the number of (correctly) transferred sheets) are determined. Furthermore, the participant receives a questionnaire to capture demographic and other experiment-related data (e.g., satisfaction with the work task and wage). Furthermore, specific questions ask whether the participants pay attention to the tax system when deciding on their labor supply, and whether participants actually calculate the effective tax rate

beforehand. All of the participants agreed to answer the questionnaire. Table 6.2 presents descriptive statistics for individual characteristics within the treatment groups.

6.4 Results

6.4.1 Tax Perception

We begin analyzing the questionnaire's data with respect to tax perception to test hypothesis 1, which states that increasing tax system complexity decreases the number of taxpayers who are able to correctly determine their tax burden. The following table summarizes the key figures from the relevant experiment-related questions and displays them both separated by treatment and in total over all of the participants.

Table 6.3: Questionnaire Statistics in Total and Separated by the Tax Systems

Tax System		A	B	C	D	Total
How pleasant did you consider the work task?	mean median SD	4.96 (5) (1.88)	4.21 (4) (1.84)	4.54 (4) (2.40)	3.75 (3) (1.80)	4.36 (4) (2.01)
Did you calculate the tax rate beforehand? (Yes)		87.5%	83.3%	58.3%	37.5%	66.7%
Statement of the calculated/ estimated tax rate (tax rate level)	mean median SD	58.33% (60%) (0.082)	57.35% (60%) (0.052)	59.06% (60%) (0.080)	63.92 (64%) (0.108)	59.53% (60%) (0.084)
To what extend did the tax influence your labor supply?	mean median SD	6.25 (6) (1.94)	5.63 (5.5) (2.58)	5.83 (6) (3.10)	6.79 (7) (2.50)	6.13 (6) (2.56)
Statement of the calculated/ estimated net wage (Cents)	mean median SD	12.78 (12) (2.15)	14.41 (14) (2.86)	13.71 (12) (5.35)	13.32 (13) (3.95)	13.56 (12) (2.87)
To what extent did the net wage influence your labor supply?	mean median SD	7.25 (7) (2.98)	6.54 (7) (2.69)	6.04 (6) (2.87)	5.88 (5) (2.72)	6.43 (7) (2.56)
How do you evaluate your tax knowledge?	mean median SD	4.50 (5) 1.72	4.04 (3.5) 2.39	4.17 (3) (2.51)	3.79 (4) (2.15)	4.13 (4) (2.19)

Note: The table presents statistical key figures of the questionnaire's experiment-related variables. The table presents the mean values of the respective variables, its median values and standard deviations (SD) separated by the Tax System. The question "How pleasant did you consider the work task?" is answered on a 10-point-scale where 1 = highly unpleasant and 10 = highly pleasant. "Did you calculate the tax rate beforehand?" denotes whether the participant claims to have calculated the tax rate beforehand. All participants are asked to state the calculated or estimated tax rate (net wage). This tax rate level (net wage) is presented by "Statement of the calculated/ estimated tax rate (net wage)" in percent (Cents). "To what extend did the tax (net wage) influence your labor supply?" is metered on a 10-point scale where 1 = no influence and 10 = high influence. "How do you evaluate your tax knowledge?" is measured on a 10-point scale where 1 = no knowledge and 10 = wide knowledge.

Table 6.3 shows that 88% of the participants in Treatment A calculated or rather noted their tax rate before the experiment. However, with increasing tax rate complexity, fewer participants did so. Thus, in Treatment D, only 38% of the participants calculated or estimated their tax rate beforehand. Hence,

all other participants did not adjust their labor supply based on the actual net wage but either used simplification rules or simply ignored taxation. Using chi-squared tests, we test whether the tax rate calculation occurs independently of the underlying tax system. The results reveal that highly significantly fewer participants in the most complex Tax System D calculate or estimate their tax rate than participants in the least complex Tax Systems A and B ($p = 0.000$ and $p = 0.001$, respectively). Even when comparing Tax System A to C (B to C), the results show that under the more complex System C, (weakly) significantly fewer participants calculate their tax rate beforehand than under Tax System A (B), with a p-value of 0.023 ($p = 0.057$). Thus, we can conclude that the number of participants who calculate or estimate their tax rate before determining their labor supply is higher in the less complex Tax Systems. Hence, we find support for hypothesis 1, which states that increasing Tax System complexity decreases the number of taxpayers who are able to correctly determine their tax burden.

Analyzing the statements on the calculated or estimated tax rates, we observe that participants in Tax System D overestimate the tax rate on average. Of the 24 participants in Tax System D, 12 indeed overestimate the actual tax burden of 60%, whereby 3 participants claim that the tax rate is 80%. This level corresponds to the sum of all of the tax rates on the tax system sheet and thus indicates that these participants used simplification rules for the calculation. We run pairwise Mann-Whitney U tests to analyze whether the difference in the stated tax rates is significant.[7] These tests reveal that, comparing Tax System A with D (B with D), participants in Tax System D (highly) significantly overestimate their tax rate in comparison with System A (B), with a p-value of 0.019 (0.005). These results additionally support hypothesis 1. Finally, the work task is perceived significantly more pleasantly by participants in Tax System A than by those in System D ($p = 0.036$). As the participants are randomly assigned to the tax systems, this result indicates a deterrent effect of tax rate complexity.

6.4.2 Accomplishment of the Work Task

We continue analyzing the experimental data with respect to the work task to test hypothesis 2 (hypothesis 3), which states that increasing tax system complexity leads to an adapted labor supply (work performance). Labor supply can be measured either by WORKING TIME or by the number of (CORRECTLY) TRANSFERRED SHEETS. Because the work task itself does not require any special abilities other than a high level of concentration, we utilize productivity and the ERROR RATE as a measure of work performance. We define productivity as the ratio of TRANSFERRED SHEETS to the WORKING TIME needed, whereby GROSS PRODUCTIVITY (NET PRODUCTIVITY) considers all (only CORRECTLY) TRANSFERRED SHEETS. The ERROR RATE is the ratio of INCORRECTLY TRANSFERRED SHEETS to all TRANSFERRED SHEETS. We inspect the work performance for

[7] The Mann-Whitney U test is a non-parametric test for dependent samples (two-tailed) that analyzes whether two independent distributions (samples) belong to the same population. Because of nature of the data, we need to use non-parametric tests; see Mann and Whitney (1947) as well as Wilcoxon (1945).

accuracy because only correctly transferred sheets are paid. The following table presents the labor supply, work performance and gross wage.

Table 6.4: Overall Labor Supply, Work Performance and Gross Wage

	mean	median	minimum	maximum	SD	total
WORKING TIME IN MINUTES	102.84	102.50	11.00	181.00	47.27	9,873.00
TRANSFERRED SHEETS	74.97	66.50	4.00	199.00	48.09	7,197.00
CORRECTLY TRANSFERRED SHEETS	63.50	57.00	1.00	176.00	43.63	6,096.00
ERROR RATE	0.190	0.142	0.000	0.857	0.146	---
GROSS PRODUCTIVITY	0.686	0.682	0.221	1.725	0.245	---
NET PRODUCTIVITY	0.565	0.557	0.043	1.560	0.240	---
GROSS WAGE	19.05	17.10	0.30	52.80	13.09	1,828.80

Note: The table presents statistical key figures for the measures of labor supply (WORKING TIME IN MINUTES, (CORRECTLY) TRANSFERRED SHEETS) and work performance (ERROR RATE and GROSS and NET PRODUCTIVITY) as well as the payment (GROSS WAGE) over all treatments (Tax Systems). We refrain from presenting the sum of the GROSS (NET) PRODUCTIVITY because this variable mirrors the ratio of (CORRECTLY) TRANSFERRED SHEET to the WORKING TIME. The same is applied for the ERROR RATE which is the ratio of INCORRECTLY TRANSFERRED SHEETS to the total TRANSFERRED SHEETS.

Participants in all of the tax systems work for 102.84 minutes on average and transfer 74.97 (63.50) sheets (correctly). The ERROR RATE is 18.97% on average. The GROSS (NET) PRODUCTIVITY, which mirrors the number of (correctly) transferred sheets per minute, is 0.686 (0.565). However, to answer the research question regarding whether tax complexity affects the labor supply and work performance, we must to extend our analysis to the treatment level.

6.4.3 The Effect of Tax Complexity on Labor Supply and Work Performance

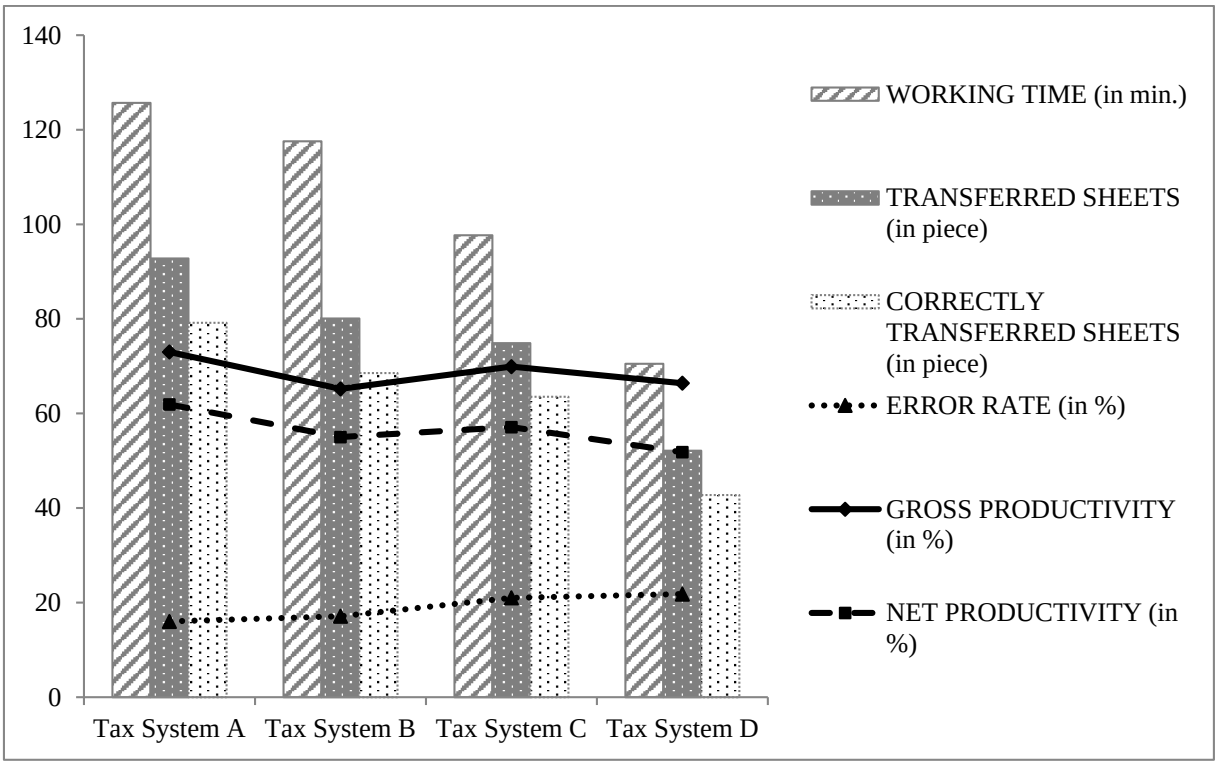
The results presented in Table 6.4 are now separately analyzed for each tax system. Thus, Table 6.5 allows an initial impression of tax complexity effects on labor supply and performance. Analyzing the research question, the main result from Table 6.5 is that participants work longer and transfer more sheets the less complex the underlying tax system is. Participants in the least complex tax system (System A) work the longest and transfer the most sheets (correctly) on average. Additionally, the average GROSS and NET PRODUCTIVITY as well as (necessarily) the GROSS WAGE are highest in Tax System A. As another measure for productivity, the ERROR RATE increases with increasing tax system complexity. These results are illustrated in Figure 6.1.

Table 6.5: Labor Supply, Work Performance and Gross Wage Separated by the Tax Systems

		mean	median	SD	minimum	maximum
A	WORKING TIME IN MINUTES	125.67	138.00	39.62	35.00	181.00
	TRANSFERRED SHEETS	92.79	80.50	41.83	16.00	196.00
	CORRECTLY TRANSFERRED SHEETS	79.17	71.50	38.74	14.00	172.00
	ERROR RATE	0.160	0.131	0.074	0.077	0.358
	GROSS PRODUCTIVITY	0.730	0.690	0.213	0.408	1.167
	NET PRODUCTIVITY	0.619	0.587	0.205	0.284	1.024
	GROSS WAGE	23.75	21.45	11.62	4.20	51.60
B	WORKING TIME IN MINUTES	117.54	113.50	38.23	54.00	171.00
	TRANSFERRED SHEETS	80.08	74.00	45.05	20.00	198.00
	CORRECTLY TRANSFERRED SHEETS	68.54	61.50	41.70	8.00	176.00
	ERROR RATE	0.171	0.119	0.131	0.061	0.680
	GROSS PRODUCTIVITY	0.652	0.625	0.233	0.333	1.329
	NET PRODUCTIVITY	0.550	0.516	0.052	0.140	1.181
	GROSS WAGE	20.56	18.45	12.51	2.40	52.80
C	WORKING TIME IN MINUTES	97.67	92.50	50.42	15.00	174.00
	TRANSFERRED SHEETS	74.88	61.00	52.41	6.00	199.00
	CORRECTLY TRANSFERRED SHEETS	63.54	51.00	46.42	1.00	169.00
	ERROR RATE	0.210	0.132	0.206	0.057	0.857
	GROSS PRODUCTIVITY	0.699	0.726	0.055	0.304	1.363
	NET PRODUCTIVITY	0.571	0.602	0.247	0.043	1.158
	GROSS WAGE	19.06	15.30	13.93	0.30	50.70
D	WORKING TIME IN MINUTES	70.50	66.00	42.05	11.00	139.00
	TRANSFERRED SHEETS	52.13	28.50	46.09	4.00	188.00
	CORRECTLY TRANSFERRED SHEETS	42.75	23.00	41.67	3.00	170.00
	ERROR RATE	0.218	0.188	0.141	0.000	0.613
	GROSS PRODUCTIVITY	0.664	0.711	0.298	0.221	1.725
	NET PRODUCTIVITY	0.518	0.477	0.275	0.221	1.560
	GROSS WAGE	12.83	6.90	12.50	0.90	51.00

Note: The table presents statistical key figures for the measures of labor supply (WORKING TIME IN MINUTES, (CORRECTLY) TRANSFERRED SHEETS) and work performance (ERROR RATE and GROSS and NET PRODUCTIVITY) as well as the payment (GROSS WAGE) separated by treatment (Tax System). GROSS (NET) PRODUCTIVITY is the ratio of (CORRECTLY) TRANSFERRED SHEET to the WORKING TIME.

Figure 6.1: Presentation of Labor Supply and Work Performance Separated by Tax System



Note: This figure illustrates the mean values of the measures for labor supply (WORKING TIME IN MINUTES, (CORRECTLY) TRANSFERRED SHEETS in piece) and work performance (ERROR RATE and GROSS and NET PRODUCTIVITY (all in %)) separated by the respective Tax System. GROSS (NET) PRODUCTIVITY mirrors the ratio of (CORRECTLY) TRANSFERRED SHEET to the WORKING TIME. ERROR RATE is the ratio of INCORRECTLY TRANSFERRED SHEETS to the number of total TRANSFERRED SHEETS.

We again run pairwise Mann-Whitney U tests to verify whether the differences in labor supply and performance between the tax systems are significant. With regard to the labor supply, we test for significant differences in WORKING TIME and (CORRECTLY) TRANSFERRED SHEETS. Furthermore, we test for significant differences in the ERROR RATE as well as in GROSS and NET PRODUCTIVITY to account for work performance. Table 6.6 first presents the differences in the respective variables between all of the tax systems whereby the initial values are displayed in Table 6.5. For each treatment comparison, the measure's value in the second tax system is subtracted by the value in the first tax system. Table 6.6 also displays the respective significance levels. Thus, we can analyze how the differences in labor supply and performance between the treatments develop with regard to the level of complexity.

Table 6.6: Pairwise Treatment Comparisons of Labor Supply and Work Performance

Compared Tax Systems	Working Time in Minutes	Transferred Sheets	Correctly Transferred Sheets	Error Rate	Gross Productivity	Net Productivity
A – B	8.125	12.708	10.625	-0.010	0.078	0.860
A – C	28.000*	17.917	15.625	-0.050	0.032	0.047
A – D	55.167***	40.667***	36.417**	-0.057	0.066	0.568**
B – C	19.875	5.208	5.000	-0.040	-0.046	-0.021
B – D	47.042***	27.958**	25.792**	-0.047*	-0.012	0.032
C – D	27.167*	22.750*	20.792*	-0.008	0.034	0.053

Note: We run Mann-Whitney U tests to make pairwise treatment comparisons of the labor supply measures (WORKING TIME and (CORRECTLY) TRANSFERRED SHEETS) and of performance measures (ERROR RATE and GROSS and NET PRODUCTIVITY). The value states the difference in the mean values of the respective measure, which are displayed in Table 6.5, whereby the value in the second tax system is subtracted by the value in the first tax system. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Regarding the labor supply, Table 6.6 demonstrates that there are significant differences between the treatments. Especially when comparing Tax Systems A and B, whose complexity is rather low, to Tax System D, which presents the highest level of complexity, participants work highly significantly longer and transfer significantly more sheets (correctly) under less complex tax systems. This result is even true when comparing System C with System D. Under the slightly less complex Tax System C, participants offer weakly significantly more labor supply. However, the level of significance decreases with increasing tax system complexity. The comparisons of System A with B, System A with C and B with C reveal almost no significant differences. Nevertheless, these results show that our hypothesis 2, whereby increasing tax system complexity leads to an adapted labor supply, is supported. To be more precise, under high tax system complexity, the labor supply is lower.

Regarding work performance, Table 6.6 shows that there are almost no significant differences in the ERROR RATE or GROSS and NET PRODUCTIVITY. Only when comparing Tax System A with System D is the difference in the NET PRODUCTIVITY significant. Under the least complex Tax System A, participants transfer significantly more sheets correctly than under the most complex Tax System D. However, we cannot establish a constantly significant impact of complexity on speed, concentration and thus productivity. This result might occur because productivity is particularly restricted by physiological and cognitive aspects that can only be indirectly influenced by the participants. Thus, we cannot support hypothesis 3, whereby increasing tax system complexity leads to adapted work performance.[8]

[8] We also run multivariate analyses to test the influence of tax complexity on work performance. Thus, we run the same OLS regressions as those for the labor supply (see Table 6.7), i.e.; we stepwise control for the tax system's impact, demographic variables and experimental characteristics. We alternatively use the ERROR RATE and GROSS and NET PRODUCTIVITY as dependent variables. We find no significant influence of tax system complexity on work performance. Thus, we refrain from further analyses in this paper.

6.4.4 Regression Analysis

In addition to the non-parametric tests, we run OLS regressions with WORKING TIME and number of TRANSFERRED SHEETS as the dependent variables.[9] Thus, in addition to the treatment effects, we can also control for the influence of demographic variables and experiment-related characteristics. The regression equation therefore becomes:

$$y_i = \beta_0 + \underbrace{\beta_1 Dummy_i^B + \beta_2 Dummy_i^C + \beta_3 Dummy_i^D}_{dummy\ variables\ for\ tax\ system\ treatments} + \underbrace{\beta_k X_{i,k}}_{control\ variables} + \varepsilon_i \quad (6.1)$$

To analyze the influence of the tax system treatments on the labor supply, we use dummy variables for each treatment. Each variable takes the value of 1 if the participant receives the respective treatment (0 otherwise). In model 1 (model 2), we first only test the tax system's influence on the WORKING TIME (TRANSFERRED SHEETS), and Tax System A serves as the reference group. Thereby, we verify the non-parametric results. Second, we extend the regressions in model 3 (model 4) using the matrix **X**, which contains individual-related control variables such as AGE, GENDER, INCOME 501 – 1,000 EURO (middle income), INCOME ABOVE 1,000 EURO" (high income) and ECONOMICS MAJOR.[10] Finally, we further extend model 3 (model 4) using experiment-related characteristics such as SATISFACTION WITH THE WORK TASK, TAX RATE CALCULATION, IMPACT OF TAXATION and IMPACT OF NET WAGE on the labor supply, perceived FAIRNESS OF THE PAYMENT and TAX KNOWLEDGE. The results of these last two regressions are displayed in model 5 (model 6). The regression results are presented in Table 6.7. The dependent variable alternates between the WORKING TIME (models 1, 3 and 5) and the number of TRANSFERRED SHEETS (models 2, 4 and 6).

[9] We also run OLS regressions with the number of CORRECTLY TRANSFERRED SHEETS as the dependent variables. The results are robust to those for the overall number of transferred sheets.

[10] The lowest income category of UP TO 500 EURO monthly disposable income serves as the reference category for the other income groups. We aggregate three income categories from the questionnaire to the highest income class, as there are only ten participants in the category of 1,000 – 1,500 EURO, no participants in the category of 1,501 – 2,000 EURO and two participants in the category of ABOVE 2,000 EURO. Thus, this aggregated category of twelve participants best reflects higher-income students.

Table 6.7: OLS Regression (Dependent Variable: WORKING TIME and TRANSFERRED SHEETS)

	model 1	model 2	model 3	model 4	model 5	model 6
Dependent Variable	WORKING TIME	SHEETS	WORKING TIME	SHEETS	WORKING TIME	SHEETS
CONSTANT	125.667***	92.792***	112.170***	89.605***	69.609*	39.132
	(8.745)	(9.493)	(30.965)	(33.415)	(36.661)	(35.853)
Treatment-Related Characteristics						
Tax System B	-8.125	-12.708	-16.440	-23.950*	-11.733	-20.030*
	(12.367)	(13.424)	(13.256)	(14.305)	(12.170)	(11.902)
Tax System C	-28.000**	-17.917	-33.057**	23.987*	-33.831***	-24.395**
	(12.367)	(13.424)	(12.734)	(13.741)	(11.950)	(11.687)
Tax System D	-55.167***	-40.667***	-57.806***	-45.058***	-53.753***	-38.406***
	(12.367)	(13.424)	(12.663)	(13.665)	(13.316)	(13.023)
Individual-Related Characteristics						
AGE			1.372	1.128	2.109	2.270*
			(1.242)	(1.340)	(1.310)	(1.281)
GENDER			-7.614	-4.607	-4.357	1.131
			(9.294)	(10.029)	(8.882)	(8.686)
INCOME 501 – 1,000 EURO			-20.768*	-27.727**	-19.721*	-27.464***
			(10.974)	(11.842)	(10.114)	(9.921)
INCOME ABOVE 1,000 EURO			-16.541	-21.362	-25.494	-30.467*
			(17.883)	(19.298)	(16.480)	(16.117)
ECONOMICS MAJOR			3.765	4.295	7.842	13.127
			(11.353)	(12.251)	(11.491)	(11.237)
Experiment-Related Characteristics						
SATISFACTION WITH WORK TASK					6.808***	8.504***
					(2.332)	(2.280)
TAX RATE CALCULATION					-24.165**	-17.749
					(11.035)	(10.792)
IMPACT OF TAXATION					-4.425**	-5.715***
					(2.053)	(2.007)
IMPACT OF NET WAGE					4.882**	4.614*
					(2.409)	(2.356)
FAIRNESS OF PAYMENT					1.784	2.826
					(2.275)	(2.225)
TAX KNOWLEDGE					-1.539	-4.689**
					(2.235)	(2.186)
Model-Characteristics						
No. of Observations	96	96	96	96	96	96
Adjusted R-square	0.179	0.095	0.177	0.074	0.338	0.388
Model Significance	0.000	0.027	0.001	0.063	0.000	0.000
Wald Tests on Coefficient Differences						
Tax System B vs. C	0.112	0.699	0.194	0.998	0.063	0.705
Tax System B vs. D	0.000	0.040	0.002	0.132	0.002	0.155
Tax System C vs. D	0.031	0.093	0.054	0.126	0.102	0.238

Note: In this table, the results of linear regression analyses are presented, whereby the dependent variable alternates between WORKING TIME and number of TRANSFERRED SHEETS (regression coefficients, robust standard errors in parentheses). To analyze the influence of the Tax Systems, we use three dummy variables. Each variable takes the value of 1 if the participant faces the respective Treatment (0 otherwise). Tax System A serves as reference group. In model 3 and 4, we control for individual-related characteristics. Therefore, the following variables are additionally included: AGE, GENDER (female = 0, male = 1), INCOME 501 – 1,000 EURO (1 if the participant reports to have 501 – 1,000 Euro at her disposal, 0 otherwise), INCOME ABOVE 1,000 EURO (1 if the participant reports to have more than 1,000 Euro at her disposal, 0 otherwise) and ECONOMICS MAJOR (1 if the subject studies economics or management, 0 otherwise). In model 5 and 6, we additionally control for experiment-related characteristics, such as SATISFACTION WITH THE WORK TASK (gives the participant's self-reported satisfaction with the experiment's work task, measured on an 10-point scale where 1 = highly unpleasant and 10 = highly pleasant), TAX RATE CALCULATION (1 if the participant claims to have calculated her tax rate beforehand, 0 otherwise), IMPACT OF TAXATION [IMPACT OF NET WAGE] (presents the participant's self-reported influence of the taxation [net wage] on her labor supply, metered on a 10-point scale where 1 = no influence and 10 = high influence) and FAIRNESS OF THE PAYMENT [TAX KNOWLEDGE] (gives the participant's self-reported perceived fairness of the payment [tax knowledge], measured on a 10-point scale where 1 = not fair [no knowledge] and 10 = very fair [wide knowledge]). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

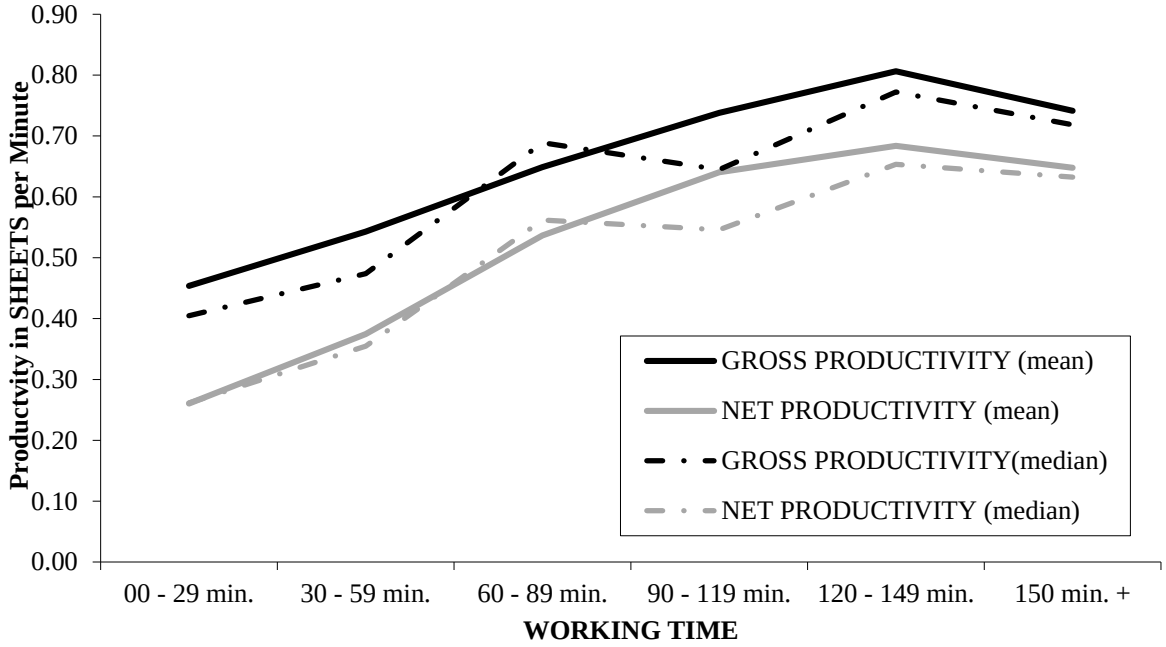
In line with the previous observations, we find a significant influence of the Tax System on the labor supply. Model 1 (model 2) shows that the WORKING TIME (number of TRANSFERRED SHEETS) is significantly lower in the most complex Tax System D compared with all three of the other Tax Systems, and the level of significance decreases with increasing tax complexity of the respective tax systems. Thus, hypothesis 2 is also supported by the OLS regression. Model 3 (model 4) analyzes whether demographic characteristics also have an influence on the labor supply. While the tax system's influence on the labor supply is verified, the results also show that participants in the middle income class work weakly significantly shorter (significantly transfer fewer sheets) than those in the low income class. This result is convincing because participants with a higher monthly disposable income might not value additional the money from the experiment as much and thus might reduce their labor supply. If they earn their monthly income themselves, they might also face higher opportunity costs. However, the results for all of the other individual-related characteristics, including the high income class, are not significant.

Model 5 (model 6) controls for whether endogenous experiment-related characteristics additionally influence labor supply decisions. Again, the influence of tax rate complexity on the labor supply is demonstrated, and the level of significance still decreases with increasing tax complexity of the respective tax systems. Furthermore, middle income participants still offer significantly less labor supply than low income participants. Analyzing the experiment-related data, we find that participants who are satisfied with the work task offer highly significantly more labor supply. This result is obvious, as participants who perceived the work task as unpleasant could either pause or leave the experiment at any time (on- or off-the-job leisure). Furthermore, the IMPACT OF TAXATION (IMPACT OF NET WAGE) is significantly negatively (positively) correlated with labor supply; i.e., the more the participants are influenced by the presented tax system (net wage), the less (more) labor supply they offer. Hence, we find evidence for an effect of taxes on the labor supply.

6.4.5 Robustness Check: Labor Supply and Productivity over Time

As presented in Table 6.4, participants transfer 0.686 (0.565) sheets per minute (correctly). Thereby, the GROSS (NET) PRODUCTIVITY strongly varies in a range between 0.221 (0.043) and 1.725 (1.560) (correct) sheets per minute. We can assume that the work task productivity mainly depends on motor skills and hand-eye coordination. If productivity strongly varies over time, both measures for the labor supply, WORKING TIME and TRANSFERRED SHEETS, will not be linearly dependent for participants who offer low labor supply and those who offer high labor supply. Thus, participants who work twice as much as others might not show double the performance, even if both participants start with the same productivity. To test for such an effect, we divide the participants by their working time into 6 groups, with each accounting for a 30-minute interval. Hence, we can control to what extent the labor performance is dependent on effects that are not monetary or tax-induced.

Figure 6.2: Gross and Net Productivity over Time



Note: This figure illustrates the mean values as well as median values of the GROSS and NET PRODUCTIVITY that are both measures for work performance in accordance to the time spent in the experiment. We divide the participants into 6 groups by their time spent in the experiment, whereas one group comprises an interval of 30 minutes. GROSS (NET) PRODUCTIVITY mirrors the ratio of (CORRECTLY) TRANSFERRED SHEET to the WORKING TIME.

Figure 6.2 presents the results, which indicate that participants who work longer demonstrate higher productivity. The reasons for this finding may be diverse, depending on the directional effects. On the one hand, the result might suggest positive learning effects that increase productivity over time. However, due to data constraints, we are not able to measure productivity over time but only the average productivity per participant over her total working time. Thus, on the other hand, the results might also indicate that participants who are more productive per se work longer because labor is more valuable to them.[11] We run a non-parametric Spearman's rank correlation test to examine whether there is a significant correlation between GROSS PRODUCTIVITY and labor supply. The test verifies that there is a significant correlation of 0.395 ($p < 0.001$). Thus, participants who are able to transfer sheets more productively spend more time working. However, productivity weakly decreases in the last interval. This result might be caused by signs of fatigue due to the high level of concentration and the monotonous work task. This assumption coincides with the participants' statements in the questionnaires, which refer to concentration problems, headaches and muscular tension.

To account for the correlation between PRODUCTIVITY and WORKING TIME, we calculated productivity-adjusted values for WORKING TIME and TRANSFERRED SHEETS. Therefore, we calculate the MEDIAN GROSS PRODUCTIVITY over all participants and adjust the WORKING

[11] In the introductions, there was a sample calculation that computed the gross wage based on the productivity of one sheet per minute. This example might have caused anchor effects if participants with a much lower productivity level became frustrated. This decreased motivation might also reduce the labor supply.

TIME and TRANSFERRED SHEETS. Hence, the ADJUSTED WORKING TIME for participant i results from:

$$\text{ADJUSTED WORKING TIME}_i = \frac{\text{TRANSFERRED SHEETS}_i}{\text{median of GROSS PRODUCTIVITY}}$$

and the adjusted number of TRANSFERRED SHEETS from:

$$\text{ADJUSTED TRANSFERRED SHEETS}_i = \text{WORKING TIME}_i \cdot \text{median of GROSS PRODUCTIVITY}.$$

The ADJUSTED WORKING TIME therefore yields the time that would have been needed by participant i to transfer her actual number of sheets if she worked at the median gross productivity level. Hence, we exclude exogenous and endogenous impacts on productivity. The ADJUSTED number of TRANSFERRED SHEETS, conversely, calculates the number of sheets a participant would have transferred in her actual working time if she worked at the median gross productivity level. Thus, we imply that participants either align their labor supply on a fixed working time or on a fixed number of transferred sheets. Therefore, we exclude the effects of working speed, tax-induced productivity differences and, thus, distortions from productivity effects. The following table presents the key statistical results from the adjusted working time and the adjusted number of transferred sheets.

Table 6.8: Adjusted Labor Supply Separated by the Tax Systems

		mean	median	SD	minimum	maximum
A	ADJUSTED WORKING TIME	136.03	118.05	61.32	23.46	287.33
	ADJUSTED TRANSFERRED SHEETS	85.72	94.14	27.03	23.88	123.47
B	ADJUSTED WORKING TIME	117.40	108.48	66.04	29.32	290.26
	ADJUSTED TRANSFERRED SHEETS	80.18	77.42	26.08	36.84	116.65
C	ADJUSTED WORKING TIME	109.76	89.42	76.83	8.80	291.73
	ADJUSTED TRANSFERRED SHEETS	66.62	63.10	34.39	10.23	118.69
D	ADJUSTED WORKING TIME	76.41	41.78	67.57	5.86	275.60
	ADJUSTED TRANSFERRED SHEETS	48.09	45.02	28.68	7.50	94.82

Note: The table presents statistical key figures for the labor supply measures (WORKING TIME in minutes and number of TRANSFERRED SHEETS) assorted by treatments (Tax Systems). The measures are adjusted by the median of the gross productivity over all participants. "SD" denotes standard deviation.

Comparing the mean values of the adjusted labor supply variables to the initial values in Table 6.5, we find almost no differences for Tax System B. In contrast to this finding, the mean values for the ADJUSTED WORKING TIME (ADJUSTED number of TRANSFERRED SHEETS) are much higher (lower) in all of the other tax systems than the mean values of the initial variables. Hence, participants in Tax Systems A, C and D needed less time to transfer the actual number of sheets and transferred more sheets than the median participant. In consideration of these adjusted variables for labor supply, we once

again run pairwise Mann-Whitney U tests. Thereby, we check whether our previous bivariate results are robust to productivity effects.

Table 6.9: Pairwise Treatment Comparisons of the Adjusted Labor Supply

Compared Tax Systems:	Adjusted Working Time	Adjusted Number Of Transferred Sheets
A - B	18.630	5.542
A - C	26.265	19.100*
A - D	59.616***	37.632***
B - C	7.635	13.558
B - D	40.986**	32.089***
C - D	33.351*	18.532*

Note: We use Mann-Whitney U tests to make pairwise treatment comparisons of the adjusted labor supply measures, i.e., the ADJUSTED WORKING TIME and the ADJUSTED TRANSFERRED SHEETS. The measures are adjusted by the median of the gross productivity over all of the participants. The respective value states the difference in the mean valued of the measures that are displayed in Table 6.8, whereby the value of the second Tax System is subtracted by the value of the first Tax System. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

The results presented in Table 6.9 only weakly differ from those in Table 6.6. While there are some variations within the absolute differences in the measure's mean values, which are reasonable with regard to the results obtained in Table 6.8, the levels of significance and the differences in signs are similar. Thus, the results are robust if individual productivity differences are taken into account by adjusting the labor supply measures for the mean gross productivity over all participants. To complete the robustness check, we run the same OLS regressions as in Table 6.7, thus applying the same stepwise integration of independent variables. However, the dependent variables are alternatingly either the ADJUSTED WORKING TIME or the ADJUSTED number of TRANSFERRED SHEETS.

Table 6.10 demonstrates that the analysis of the adjusted labor supply variables yields the same results as the OLS regression with the initial values as dependent variables. The effects of tax complexity on labor supply are also verified for the variables that are adjusted for the gross productivity of the median participant. Thus, the results are robust to possible productivity effects.

Table 6.10: OLS Regression (Dependent Variable: ADJUSTED WORKING TIME and ADJUSTED TRANSFERRED SHEETS)

	model 7	model 8	model 9	model 10	model 11	model 12
Dependent Variable	WORKING TIME	SHEETS	WORKING TIME	SHEETS	WORKING TIME	SHEETS
CONSTANT	136.030***	85.723***	131.358***	76.516***	57.366	47.483*
	(13.916)	(5.965)	(48.985)	(21.123)	(52.559)	(25.008)
Treatment-Related Characteristics						
Tax System B	-18.630	-5.542	-35.110*	-11.214	-29.363*	-8.004
	(19.680)	(8.436)	(20.971)	(9.042)	(17.447)	(8.301)
Tax System C	-26.265	-19.100**	-35.164*	-22.550**	-35.762**	-23.078***
	(19.680)	(8.436)	(20.144)	(8.686)	(17.133)	(8.152)
Tax System D	-59.616***	-37.632***	-66.054***	-39.432***	-56.301***	-36.667***
	(19.680)	(8.436)	(20.032)	(8.638)	(19.091)	(9.083)
Individual-Related Characteristics						
AGE			1.654	0.936	3.328*	1.439
			(1.964)	(0.847)	(1.878)	(0.894)
GENDER			-6.754	-5.194	1.658	-2.972
			(14.703)	(6.340)	(12.734)	(6.059)
INCOME 501 – 1,000 EURO			-40.647**	-14.167*	-40.261***	-13.453*
			(17.360)	(7.486)	(14.544)	(6.920)
INCOME ABOVE 1,000 EURO			-31.316	-11.284	-44.663*	-17.390
			(28.290)	(12.199)	(23.628)	(11.242)
ECONOMICS MAJOR			6.296	2.568	19.244	5.349
			(17.960)	(7.744)	(16.474)	(7.838)
Experiment-Related Characteristics						
SATISFACTION WITH WORK TASK					12.467***	4.644***
					(3.343)	(1.591)
TAX RATE CALCULATION					-26.020	-16.484**
					(15.821)	(7.528)
IMPACT OF TAXATION					-8.378***	-3.019**
					(2.943)	(1.400)
IMPACT OF NET WAGE					6.764*	3.330**
					(3.454)	(1.643)
FAIRNESS OF PAYMENT					4.143	1.217
					(3.261)	(1.557)
TAX KNOWLEDGE					-6.874**	-1.050
					(3.205)	(1.525)
Model-Characteristics						
No. of Observations	96	96	96	96	96	96
Adjusted R-square	0.065	0.179	0.074	0.177	0.388	0.338
Model Significance	0.027	0.000	0.063	0.001	0.000	0.000
Wald tests on Coefficient Differences						
Tax System B vs. C	0.699	0.112	0.998	0.194	0.705	0.063
Tax System B vs. D	0.040	0.000	0.132	0.002	0.155	0.002
Tax System C vs. D	0.094	0.031	0.126	0.005	0.238	0.102

Note: In this table, the results of linear regression analyses are presented, whereby the dependent variable alternates between the ADJUSTED WORKING TIME and the ADJUSTED number of TRANSFERRED SHEETS (regression coefficients, robust standard errors in parentheses). To analyze the influence of the Tax System, we use three dummy variables. Each variable takes the value of 1 if the participant faces the respective treatment (0 otherwise). Tax System A serves as reference group. In model 9 and 10, we control for individual-related characteristics. Therefore, the following variables are additionally included: AGE, GENDER (female = 0, male = 1), INCOME 501 – 1,000 EURO (1 if the participant reports to have 501 – 1,000 Euro at her disposal, 0 otherwise), INCOME ABOVE 1,000 EURO (1 if the participant reports to have more than 1,000 Euro at her disposal, 0 otherwise) and ECONOMICS MAJOR (1 if the subject studies economics or management, 0 otherwise). In model 11 and 12, we additionally control for experiment-related characteristics, such as SATISFACTION WITH THE WORK TASK (gives the participant's self-reported satisfaction with the experiment's work task, measured on an 10-point scale where 1 = highly unpleasant and 10 = highly pleasant), TAX RATE CALCULATION (1 if the participant claims to have calculated her tax rate beforehand, 0 otherwise), IMPACT OF TAXATION [IMPACT OF NET WAGE] (presents the participant's self-reported influence of the taxation [net wage] on her labor supply, metered on a 10-point scale where 1 = no influence and 10 = high influence) and FAIRNESS OF THE PAYMENT [TAX KNOWLEDGE] (gives the participant's self-reported perceived fairness of the payment [tax knowledge], measured on a 10-point scale where 1 = not fair [no knowledge] and 10 = very fair [wide knowledge]). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

6.5 Interpretation and Discussion

By conducting an experiment, we analyze the effect of tax rate complexity on the participants' labor supply. Keeping the effective tax rate and gross and net wage equal in four different tax tariff systems, we model tax tariff complexity using an increasing number of tax rates and an increasing mathematical interaction of the rates. Our results reveal that increasing tax tariff complexity drives taxpayers to decrease their labor supply, but not their work performance. We also control for exogenous differences in productivity.

In the most complex tax system, we present four tax rates that are combined additively, multiplicatively and by the deductibility of one tax for another. Here, the effective tax rate could only be calculated with great effort. Our results reveal that subjects are not able or willing to calculate the effective tax rate. Thus, they might either disregard taxes completely or use heuristics or simplification rules to adjust their labor supply. As the labor supply in the most complex tax tariff treatment is significantly lower than that in the less complex treatments, we can conclude that the subject's uncertainty regarding the actual tax payment leads to an overestimation of the tax burden. These results are verified by the questionnaire results. Thus, labor is perceived to be less valuable. Therefore, in accordance with the prior literature, the overestimation of the tax rate leads to a decrease in the labor supply.

Alternatively, the reduced labor supply may also be caused by participants' general dissatisfaction due to the many interacting tax rates. Thus, work motivation may decrease without even calculating or estimating the tax rate. As not only the tax rate itself but also the level of tax tariff complexity leads to an adjustment of the labor supply, policymakers should take tax complexity into account in tax reform discussions. They should not try to hide tax burdens behind complex tax systems to cushion the perceived burden as tariff complexity also reduces the labor supply. In fact, a more simple presentation of the underlying tax system may provide a labor incentive and decrease disincentives due to complexity.

It is apparent that not only the tax rate itself but also the level of tax tariff complexity has an influence on labor-leisure decisions. This result can mainly be found when comparing tax systems with the highest level of tax complexity difference. Analyzing the labor supply for a lower level of tax complexity difference, e.g., comparing tax systems A and B, we find no (consistent) significant differences. However, the general assumption of a precisely calculating taxpayer in theoretical models is questionable if these models claim to predict actual behavior, especially given the background involving the present tax tariff complexity. Subjects use heuristics or simplification rules that decisively influenced their behavior. These heuristics do not drive subjects to orientate on pre-tax values, which would have caused an increased labor supply; rather, they overestimate their tax burden and thus decrease their labor supply.

6.6 Limitations and Future Research

Despite our findings, we want to stress some limitations that should be considered when interpreting or generalizing our results. We conduct the experiment with a rather small number of participants, i.e., 24 participants per tax system (96 in total). Thus, this study should be conducted with a higher number of participants to verify our results. Although students are accepted as a good measure of people's general behavior, they present a rather homogenous group when considering age and education. Thus, the participants' variety should be extended in future research to extend its representativeness.

Furthermore, we refrain from modelling learning effects. Learning effects can occur either with respect to productivity if participants become used to the work task over time or with respect to the tax system. If participants are continuously confronted with taxes and their complexity in practice, they might decrease the usage of heuristics or simplification rules for determining their tax burden. However, as the income tax system faces an ongoing flux (there are constantly new tariff norms in Germany), learning effects appear to play only a minor role within the concept of our research question. If taxpayers bother to calculate their tax burden ex ante in reality to make their decision, it is rather unrealistic that the same calculations will be possible in future years if the tax system experiences changes. Nevertheless, we cannot strictly disregard learning effects, so future research might take them into consideration.

We additionally want to raise awareness of the tax labelling issue. As already presented in the description of the experimental design, we try to eliminate labelling or affectation effects through the use of real terms when labelling the different tax rates in each treatment. Thereby we refrain from abstract terminology, such as tax 1, tax 2 etc., so that we do not create an absolutely theoretical framework. Furthermore, we intended to name taxes similarly that work mathematically equally, which we did successfully but with one exception. Nonetheless, it is arguable whether the different labelling of taxes between the treatments may have caused confounders which cannot be measured. As the most complex tax system is also modelled by the highest number of interacting, differently labelled tax rates, we cannot rule out whether the tax rates' labels also affected participants' labor supply. Thus, future research should control for labelling effects.

Overall, our research represents another step toward establishing experimental tax research in Germany and should be extended in future studies. Additionally, we also contribute to theoretical research on tax policy, as we stress the behavioral relevance of complexity, which should also be considered in future models and within tax reform discussions. Above all, the thesis that policymakers can hide fees or taxes behind complex tax systems to cushion tax resistance cannot be confirmed within our experiment.

6.7 Appendix A: Experimental Instructions

Below is the experiment's instruction. The experiment was conducted in German, hence, everything shown below was translated into English.

Instruction

By participating in this experiment you have the opportunity to earn money. The amount of money at the end of the experiment depends on your personal effort. Please read these instructions carefully. If you have any questions, please contact the experimenter.

The aim of this experiment is to obtain information about your personal labor supply. For this purpose you will be confronted with a real work situation where you can earn money. In order to compare the data of different participants, the work task is an activity that does not require any previous knowledge or special skills. We would like to point out that conversations with other participants, leaving the site and using the computer for private purposes during the experiment are not allowed. After you have read these instructions, you will receive several solution sheets of an exam, which have to be digitized. For this purpose you will find an Excel file on your computer into which the solution sheets are to be transferred. After reading these instructions, the experimenter will show you how the digitization procedure is to be carried out.

Depending on the number of correctly digitized sheets, you will receive your wage payment at the end of the experiment. For each digitized sheet you will receive a ***gross wage of 30 cents****. For example, if you digitize one sheet per minute on average, you will receive a wage of 18 euros gross per hour; if you digitize 1.5 sheets per minute on average, you will receive a wage of 27 euros gross per hour.* Please note: You still have to pay taxes on your gross wage, only the rest (net wage) will be paid to you in cash at the end of the experiment. The amount of taxes to be paid or the corresponding tax rate can be found on the tax rate sheet, which will be handed out separately. Please read it carefully before starting the activity!

You decide yourself about your labor supply! This means that you can work as long and as much as you want. So you can stop the experiment at any time and you will receive your wage depending on the work you have done. Please contact the experimenter if you want to end the experiment.

Have fun!

6.8 Appendix B: Tax System Sheets

The tax system sheets were standardized to the same size when they were distributed to the participants. The original tax system sheets were written in German, thus everything shown below was translated into English.

A

(1) The tax on the income amounts to 60%.

B

(1) The tax on the income (income tax) amounts to 40%.

(2) Additionally, there is an additive tax on the income amounting to 20%.

C

(1) The tax on the income (income tax) amounts to 40%.

(2) Additionally, there is a surcharge tax on the income tax (not on the income itself!) amounting to 20%.

(3) Furthermore, there is a state tax on the income amounting to 12%.

D

(1) The tax on the income (income tax) amounts to 20%.

(2) Additionally, there is a surcharge tax on the income tax (not on the income itself!) amounting to 20%.

(3) Furthermore, there is a state tax on the income amounting to 20%.

(4) Additionally, there is a community tax on the income also amounting to 20%.

(5) The community tax decreases the level of income on which the income tax is levied (deductibility of the community tax for income tax' calculations).

6.9 Appendix C: Questionnaire

Below is the questionnaire from the experiment. The questionnaire was also conducted with z-tree and in German, hence, everything shown below was translated into English.

How pleasant did you consider the work task?

Highly unpleasant ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Highly pleasant

Did you calculate or try to calculate the tax rate beforehand?

☐ Yes
☐ No

Please tell us the effective (total) tax rate applicable to you now once again, or please try to calculate it now, if you have not already done so during the experiment. (in Percent)

☐

To what extend did the tax influence your labor supply?

no influence ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ high influence

Did you calculate your net wage beforehand?

☐ Yes
☐ No

To what extent did the net wage influence your labor supply?

no influence ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ high influence

How fair did you perceive the remuneration for the job you were asked to do?

not fair at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ very fair

How would you rate your own tax law knowledge?

no knowledge ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ above average knowledge

Which study program are you enrolled for? ☐

Which academic semester are you in? ☐

Have you already taken courses in the area of specialization "Taxation" during your studies?

☐ Yes
☐ No

Gender:

☐ Female
☐ Male

Age (in years): ☐

What is your monthly disposable income?

☐ < 500 €
☐ 501 € - 1,000 €
☐ 1,001 € - 1,500 €
☐ 1,501 € - 2,000 €
☐ > 2,000 €

Chapter 7

Mental Accounting and the Timing of Pension Taxation*

Abstract

We study whether the timing of pension taxation influences work effort and risk-taking. In a real effort experiment, participants first earn money and then invest the money earned. Participants in the immediate (deferred) taxation treatment (don't) pay taxes on their wages, but their invested pension capital and the respective returns are tax-exempt (taxed). After-tax payoffs are equal in the immediate and deferred taxation systems. However, we find that participants in the deferred taxation treatment perceive their wage as significantly more fair, and this perception indirectly increases work effort. Moreover, we find that deferred taxation decreases risk-taking.

Keywords: Deferred Taxation · Pension Tax Reliefs · Mental Accounting · Tax Perception · Behavioral Taxation

JEL Codes: H24 · H31

* This chapter is co-authored by Prof. Dr. Kay Blaufus (Leibniz University Hannover), Prof Dr. Jochen Hundsdoerfer (Freie University Berlin) and Dr. Matthias Sünwoldt (Freie University Berlin).
We would like to thank Ido Erev, Martin Fochmann, Martin Jacob, Florian Morath, and the participants at the 2015 Conference of the German Association of Experimental Economics, the 2015 Closing Workshop on "Coping with Difficult Decisions - An Experimental Economics Perspective," the 2016 WHU Accounting & Tax Research Seminar, the 2016 ZEW Workshop on "Lab Experiments in Public Economics", the 2016 VHB Conference in Munich, and the 2016 EAA Conference in Maastricht.

7.1 Introduction

Does the timing of taxation affect the behavior of individuals even if it does not affect their wealth? We answer this question affirmatively. As a result of mental accounting, the timing of taxation is important even in the absence of wealth differences between alternatives.

While this question is of general interest, it is especially important in the field of pension taxation. Most OECD countries use tax incentives to encourage private retirement savings plans (Yoo and De Serres, 2005). Some countries do not tax interests earned in savings plans funded with after-tax contributions, whereas other countries defer payroll taxation on contributions until pensions are paid out. The first method implies immediate taxation of wages but no taxation of saving returns (TEE, i.e., taxable contributions, exempt accumulations and exempt withdrawals). This method has been implemented in countries such as Hungary, Luxembourg, and the U.S. (Roth individual retirement arrangements and Roth 401(k) plans) and has recently been proposed in the UK (HM Treasury, 2015). The second method implies deferred taxation of wages and savings returns (EET, i.e., exempt contributions, exempt accumulations and taxable withdrawals), and it is currently offered in countries such as Austria, the Netherlands, Norway, the UK, and the U.S. (401(k) plans).

We concentrate on the question of whether the timing of taxation (immediate taxation or TEE vs. deferred taxation or EET) affects individual labor supply and investment risk decisions. Suppose that a subject earns € 100 and wants to invest in a one-year pension plan. Under an immediate taxation scheme (TEE), the subject pays taxes on her wage income but does not pay taxes on future withdrawals, including interest income. Assuming an interest rate of 10% and a tax rate of 60%, a total of € 40 remains for her investment, and she receives € 44 after one year. Under deferred taxation (EET), wage income is tax free (i.e., contributions are tax deductible), but withdrawals from the pension plan are fully taxed. In this case, investing € 100 in the pension plan leads to a withdrawal of € 110 after one year and € 44 after taxes. Thus, the timing of taxation (immediate or deferred) does not affect her wealth. From a neoclassical perspective, given a time-constant tax rate, the timing of taxation should not affect individual behavior because both alternatives are equivalent in present value terms.

This "neutrality result" has significantly influenced tax research and the evaluation of tax policy options. This view has been considered in the discussion of different pension tax systems (e.g., Yoo and De Serres, 2005; Huang, 2008). Furthermore, it is used to study tax effects in a variety of contexts, such as the choice between present and future consumption, between lifetime and testamentary gifts, between retention and distribution of corporate earnings, between receiving or deferring income, and between different forms of doing business (Warren, 1986). Despite the importance of this equivalence between immediate and deferred taxation for tax policy and research, its empirical validity is an open issue. Only recently has a direct test been offered by Beshears et al. (2017). Their results contradict the neoclassical neutrality assumption regarding the timing of taxation, as they find that future after-tax pensions will differ between immediate and deferred pension tax systems. Their findings suggest that taxpayers

generally neglect taxes and thus do not adjust their contributions to pension plans according to different tax treatments. This behavior in turn leads to different expected pensions in immediate and deferred taxation systems. The current paper adds to the discussion of the economic equivalence of immediate and deferred pension taxation by studying whether the timing of taxation influences subjects' work effort and risk-taking decisions. Moreover, we provide an additional behavioral explanation why immediate and deferred pension tax systems could lead to divergent behavioral responses.

The equivalence of deferred and immediate taxation is based on the assumption that subjects accurately estimate their total tax burden. However, increasing evidence suggests that many subjects misperceive taxes because they do not pay attention to less salient taxes (e.g., Sausgruber and Tyran, 2005; Chetty et al., 2009; Blumkin et al., 2012; Fochmann and Weimann, 2013), because they focus on pre-tax-values instead of after-tax returns (Fochmann et al., 2013; Weber and Schram, 2017), because they use simple decision heuristics (Blaufus et al., 2013) or because taxes induce negative emotions that affect decision making (Blaufus and Möhlmann, 2014).

With respect to the equivalence between immediate and deferred taxation, we propose that mental accounting (Thaler, 1985) leads to a deviation from the "neutrality result." Prior research provides evidence that mental accounting is of relevance in the tax context: A tax refund delivered in monthly amounts increases current spending more than the same yearly total tax reduction delivered in one lump sum (Chambers and Spencer, 2008), and a tax decrease implemented gradually over several years leads to a greater increase in risky investment than a tax change implemented all at once (Falsetta et al., 2013). In the context of life-cycle consumption, Thaler (1990) argues that individuals tend to use mental accounts when evaluating savings and consumptions. He considers three broad accounts when categorizing types of wealth: a current income account, an asset account and a future income account. Based on the fungibility of money, the marginal propensity to consume should be equal for all three accounts. These mental accounts could explain some of the observed saving and consumption anomalies, such as the tendency for consumption to respond too sensitively to current income. Individuals seem to evaluate the different mental accounts separately. In this context, Feldman (2010) analyses the 1992 U.S. income tax withholding decrease in which the marginal tax rate was kept constant. Hence, taxpayers faced lower monthly payments while facing a higher lump-sum at the year-end. Feldman (2010) finds that this income shift from one mental account (wealth mental account) to another (consumption mental account) decreases the probability that households invest to a tax-preferred retirement account. Considering Thaler's concept of mental accounting, we expect that some subjects assign their wealth decisions to two separate mental accounts (a work account and an investment account) rather than determining their wealth based on one aggregated decision. If these individuals are subject to deferred (immediate) taxation, they will make work decisions as if wages are tax free (subject to tax) and make investment decisions as if returns are fully taxable (tax exempt). Thus, we expect that work and investment decisions differ between immediate and deferred taxation.

Despite the usual concerns about external validity, a laboratory experiment has obvious advantages in answering our research question. First, the neutrality between deferred and immediate pension taxation requires that present and future tax rates are equal and that future tax rates are known with certainty. This requirement can easily be met only in an experimental situation. Second, productivity and work performance often may be only roughly estimated using archival data. Hence, to test the hypotheses, we conduct a real-effort laboratory experiment with subjects randomly assigned to either an immediate or Deferred Taxation Treatment. In the first step, subjects perform a work task lasting one hour, and we measure their work effort. In the second step, subjects must invest their earned money. In this way, we simulate a simple, compulsory pension contribution system. In the investment phase subjects make five lottery decisions, and in each decision, they choose between two lotteries that differ in risk but not in expected returns.

We find that mental accounting is important. In line with the mental accounting hypothesis, subjects in the Deferred Taxation Treatment perceive their wage as significantly more fair, and this perception (indirectly) increases working effort. Moreover, subjects in the Deferred Taxation Treatment make less risky investments. Thus, the presumed neutrality regarding the timing of taxation does not hold. This finding is of relevance to the current debate in countries such as the United Kingdom regarding a change from deferred pension to immediate pension taxation. Policy makers who decide between deferred and immediate taxation should consider that neoclassical predictions could be misleading. Whereas neoclassical economics would predict the same tax revenues (in present value terms), our results suggest the tax revenues differ because of the different behavioral effects of immediate and deferred taxation. In addition to observing direct effects on after-tax pensions (Beshears et al., 2017), we demonstrate that a change may cause unexpected effects on work effort and risk-taking.

The remainder of this article is organized as follows. In the next section, we derive our hypotheses. In the third section, we describe the experimental protocol and sample. The results are provided and discussed in Section 7.4. We present robustness tests in Section 7.5 and conclude the paper in Section 7.6.

7.2 Hypothesis Development

7.2.1 Work Effort

Given a world where W denotes the wage that is used for saving purposes, i represents the market interest rate that applies to the savings over period n; τ denotes the investor's tax rate, which is constant over time; and the investor's wealth WE_i under immediate taxation is given by the following equation:

$$WE_i = W(1-\tau)(1+i)^n. \tag{7.1}$$

Under immediate taxation, the wage income that is used for savings is subject to taxes. However, the return from these savings is tax exempt. In contrast, under deferred taxation, the wage income that is

used for savings remains tax free, but the return from savings is fully taxable. Thus, deferred taxation leads to the following equation for the investor's wealth:

$$WE_d = W(1+i)^n(1-\tau). \tag{7.2}$$

Both tax systems lead to the same wealth if we assume equal wages, the same time horizon, and the same tax and interest rates ($WE_i = WE_d$). Thus, according to neoclassical economics, working and savings behavior are unaffected by the timing of taxation. An overview of the two tax systems is presented in Table 7.1.

Table 7.1: Overview of the Taxation System

	Deferred Taxation	**Immediate Taxation**
Work stage	Work: *no tax*	Work: *taxation*
Investment stage	Investment: *taxation*	Investment: *no tax*
	Equal present values	

Mental accounting could distort this wealth neutrality. The concept of mental accounts is based on the idea that individuals evaluate financial activities comparable to the managerial and financial accounting that is used by firms and other organizations (Thaler, 1985, 1999). Mental accounting consists of three main components (Thaler, 1999): perception of outcomes and evaluation of decisions, frequency of evaluation or "balancing", and assignment of financial activities. The last component is particularly relevant for our experimental setting. In mental accounting, similar to financial accounting, individuals label expenditures and funds as different accounts based on sources and uses (e.g., regular income vs. windfall). As a consequence, the fungibility of money, introduced by the account-based budgeting, could be violated (for experimental evidence, see Heath and Soll, 1996). In our experimental design, we do not directly refer to the fungibility problem, but we argue that subjects in our experiment use two different accounts (a work income account and an investment income account) because we assume that subjects separate their income based on income sources (work income vs. investment income). If we consider that subjects use these different mental accounts, the above derived "neutrality result" may no longer hold. Under deferred taxation, subjects might perceive their wages as tax free because they neglect the indirect, less salient deferred taxation. In contrast, under immediate taxation, subjects might perceive the wage income as fully taxable, neglecting the tax advantage of tax-exempt investment income. Thus, subjects who use different mental accounts for work and investment income may perceive the wage rate to be higher under deferred taxation than under immediate taxation.

A higher perceived wage rate should lead to an increase in work effort[1] (see, for example, Fehr and Goette, 2007) as leisure becomes less valuable (*perceived substitution effect*). Thus, we would expect participants in the Deferred Taxation Treatment to work harder than those in the immediate taxation treatment. However, if subjects use mental accounting, a *perceived income effect* might also occur, and reduced effort will be the result. This *perceived income effect* will arise as subjects earn more money to invest in goods, such as leisure. Hence, the higher perceived wage rate under deferred taxation will cause a higher demand for leisure. Thus, either the opposing effects, the *perceived income* and *perceived substitution effects*, offset one another, or one predominates the other. Following the results of Fehr and Goette (2007) as well as of Sutter and Weck-Hannemann (2003) who find an inverse relationship of tax rates we assume that work effort decreases with rising perceived tax rates so that the perceived substitution effect offsets the perceived income effect. We formulate our first hypothesis accordingly:

Hypothesis 1: Work effort is higher under the deferred taxation of wages.

If we actually find that work effort is higher under deferred taxation, this result would contradict the neoclassical neutrality assumption and support the mental accounting hypothesis. By contrast, if we do not find any difference in work effort, this does not necessarily imply that subjects make decisions in line with neoclassical assumptions. Alternatively, subjects could use mental accounting while perceived income and substitution effects offset one another.

To directly test the mental accounting hypothesis, we analyze whether subjects perceive their wage rate differently depending on the taxation system. According to neoclassical theory subjects should realize that their future wealth is independent of the taxation system. Hence, subjects in the Deferred Taxation Treatment should comprehend that their wage is indirectly taxed when they have to pay taxes on their invested amount. Therefore, subjects should perceive their wage as equally high and fair under both treatments. However, if subjects actually use different mental accounts for wage and investment income they should perceive the wage rate in the immediate taxation system as less fair than the wage rate under deferred taxation. This difference results from the work mental accounting that occurs during the work task. If subjects think only within their work mental account, they encounter a net wage rate in the immediate taxation treatment but a gross rate in the Deferred Taxation Treatment. Thus, subjects should perceive the wage rate to be higher under deferred taxation only if they use mental accounting. We formulate Hypothesis 2a accordingly:

Hypothesis 2a: The perceived fairness of wages is higher under the deferred taxation of wages.

The impact of perceptions of a payment's fairness on work effort has been broadly examined in the literature. In the context of gift exchanges and reciprocity, Akerlof (1982), Akerlof (1984), Fehr et al.

[1] In our experiment, subjects decide not on the hours to work but on the effort to expend. As subjects decide between work effort and on-the-job leisure (Dickinson, 1999), we may only measure work effort within that time span.

(1998), Riedl and Tyran (2005) and Dohmen et al. (2009) find that a payment's fairness is positively correlated with work effort. Hence, the fairer subjects feel a payment is, the more effort they will exert in their labor. Thus, we formulate the Hypothesis 2b as follows:

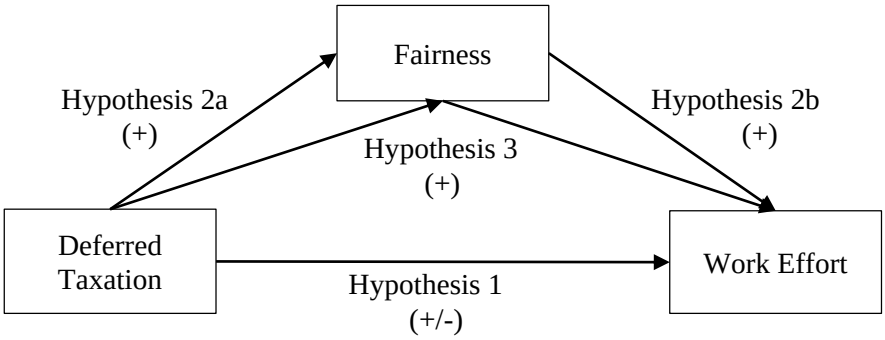
Hypothesis 2b: *The higher the perceived fairness of wages is, the greater the work effort is.*

Summarizing Hypotheses 2a and 2b, we expect participants in the Deferred Taxation Treatment to engage in mental accounting and perceive their work task payment fairer and therefore offer greater work effort. Thus, if we consider perceived fairness to be a mediator variable, we can specify our first hypothesis with regard to an indirect effect:

Hypothesis 3: *Deferred taxation indirectly increases work effort because the perceived fairness of wages operates as a mediator variable.*

The general theoretical model underpinning the testing of these hypotheses is illustrated in Figure 7.1. We expect to find a direct effect of the deferred taxation system on work effort (Hypothesis 1). Moreover, fairness mediates the relationship between the tax system and work effort. Under the deferred taxation system, the perceived fairness of a payment is higher than the perceived fairness under the immediate taxation system (Hypothesis 2a). Fairness and work effort are positively correlated: an increase in perceived fairness will improve work effort (Hypothesis 2b). Thus, we expect to find a positive indirect effect of deferred taxation on work effort (Hypothesis 3).

Figure 7.1: Theoretical Model of Fairness Mediating the Relationship between Work Effort and the Tax System



7.2.2 Risk-Taking

Rational choice theory would also predict no differences in subjects' investment decisions.[2] However, mental accounting entails a perception of full taxation (tax exemption) of investment returns under deferred (immediate) taxation, as subjects consider only their mental investment account. Thus, we again

[2] Note that the taxation of investment returns and the amount invested in the Deferred Taxation Treatment are comparable with a wealth tax in the classical framework of Atkinson and Stiglitz (1980): future wealth is reduced because of taxation, and this leads to less risk-taking than in a situation without taxation (assuming decreasing relative risk aversion). However, in contrast to Atkinson and Stiglitz (1980), we compare the deferred taxation system with an immediate taxation system, both of which offer identical future wealth.

expect differences between the two tax systems due to mental accounting. First, subjects in the deferred taxation system have more to invest because their income has not yet been taxed. This perceived *income effect* may lead to higher or lower levels of risk-taking or may not affect risk-taking at all depending on the subjects' risk preferences. If one assumes constant relative risk aversion on average (in line with the experimental results of Chiappori and Paiella, 2011), one would expect no effect. However, if one assumes decreasing relative risk aversion (in line with many experimental and empirical studies, e.g., Levy, 1994; Calvet and Sodini, 2014), one would expect greater risk-taking under deferred taxation.

Second, both investment alternatives differ solely in the variance of returns. Whereas immediate taxes have already reduced the investment amount, taxes in the Deferred Taxation Treatment reduce the variance of returns. Although after-tax payoffs are equal, the risk in the Deferred Taxation Treatment might be perceived as lower, which could increase risk-taking under deferred taxation.

Third, prior behavioral research finds that taxation results in lower risk-taking because of the additional complexity induced by taxes (Ackermann et al., 2013). If subjects engage in mental accounting, they will consider only their investment account when making portfolio decisions. Regarding the mental investment account, only subjects under deferred taxation face taxes at the investment stage, as the returns and invested capital for subjects under immediate taxation are tax exempt. Thus, the taxation of returns and invested capital in the Deferred Taxation Treatment might cause an increase in complexity, which would lower risk-taking behavior.

In sum, while a rational individual will not change her risk-taking behavior between the treatments, mental accounting leads to a difference in risk-taking behavior between deferred and immediate taxation. However, the theoretical direction is unclear and is thus an empirical question. We therefore test the following hypothesis:

Hypothesis 4: Levels of risk-taking differ between deferred and immediate taxation.

7.3 Experimental protocol and sample

7.3.1 Experimental Protocol

We use a between-subjects design with the timing of taxation (immediate versus deferred) as the treatment variable. Subjects are randomly assigned to the two treatment groups. The experiment was conducted in eight sessions at the computerized experimental laboratory of the Leibniz University Hannover. The z-Tree software was used (Fischbacher, 2007). We present a translation of the instructions in Section 7.7 (Appendix A) and screenshots of the experiment in Section 7.8 (Appendix B).

The experiment is divided into two parts: a work task and portfolio decisions. To avoid that subjects use individual scripts when interpreting loaded terms, we do not frame the experiment as a pension tax task, but use neutral language (Alm, 1991, 2010). The work task lasts one hour. During this time, the

participants are asked to digitize answer sheets for a multiple choice exam. Each answer sheet consists of 60 rows (questions) with 6 possible columns (answers). The participants are asked to correctly transfer each checked box by clicking the respective check box on the computer screen. This work task offers two advantages: On the one hand, it is largely independent of the participants' education and abilities; thus, all participants have the opportunity to earn the same money. On the other hand, the correctness of the digitized sheets is automatically controlled, thus enabling the payment to be dependent on the participants' accuracy. The participants are granted one practice period to familiarize themselves with the work's design and task. However, the practice period is not relevant to the pay-out.

Before each digitalization, the participants enter a four-digit number that identifies a certain answering scheme. We use twelve different answering schemes, each comprising 60 different four-digit numbers. Based on the answering sheet's number, the computer checks the correctness of the work task. Only accurately transferred answering sheets are paid. After each sheet, the participant is told about the correctness of her work and the amount of money she has already earned. For each correct sheet, the participant earns € 2 (before taxes).[3] A countdown projected on the front wall informs the participants how much time is left for work. During the work task, the participants are allowed to surf the internet. Thus, we offer them an on-the-job leisure alternative if they want to pause or end their work task (Dickinson, 1999).

The two treatments differ only regarding the time of taxation.[4] In the Immediate Taxation Treatment, the participants are told that their gross wage is taxed at a rate of 60%.[5]Thus, they earn a net wage of € 0.80 per correctly transferred sheet. However, their subsequent portfolio decisions have tax-free returns. In the Deferred Taxation Treatment, the participants are told that their wage is tax free but that they must pay taxes at a rate of 60% on both the returns of their portfolio decisions and the invested capital. This summarizing taxation information is presented in the first paragraph of the second part of the instructions (see Appendix A1.2). The taxation information is then given again for the working and investment periods separately. To ensure that the salience regarding taxation is the same between treatments, we include the same type of information. More precisely, we inform the participants on the taxation of their wage (tax free versus 60% tax rate on gross wages) and their investment (tax free versus 60% tax rate on invested capital and returns). Thus, both treatments lead to the same tax burden and after-tax payoff, but they differ in the timing of taxation. After the work task ends, the investment phase starts. At the beginning of the investment phase, the participants are given an overview of their work results (i.e., how many correct sheets they have entered). The Immediate Taxation Treatment group is also informed about the gross and net wage (in euros) and about withholding taxes, whereas the Deferred

[3] Tax revenues from this experiment are not distributed among participants. Instead, they are used for further experimental research at the experimental laboratory of the Leibniz University Hannover.

[4] Note that all subjects are informed about both parts of the experiment (work and investment task) and the respecting tax treatments before the experiment starts (see Figure 7.2 and Section 7.7 (Appendix A) for the instructions).

[5] We decided for this rather high tax rate to increase the tax salience.

Taxation Treatment group only receives additional information on their wage, as taxes are not withheld at that time.

In the investment phase, the participants successively confront five portfolio decisions. Due to the prior work task, we prevent a potential *house money effect* that might increase the participants' risk-taking if they do invest money that occurred as a windfall.[6] In each of the decisions, the participant is presented with two alternative independent lotteries, with one lottery always riskier than the other, but both have the same expected value. The participants in the Immediate Taxation Treatment (Deferred Taxation Treatment) are asked to distribute their entire net wage (wage) between the two lotteries in each portfolio decision. Within each lottery, three different states may occur with a probability of one-third. The lotteries are presented in Table 7.2 with the riskier lottery on the left side.

Table 7.2: Overview of Portfolio Decision Lotteries

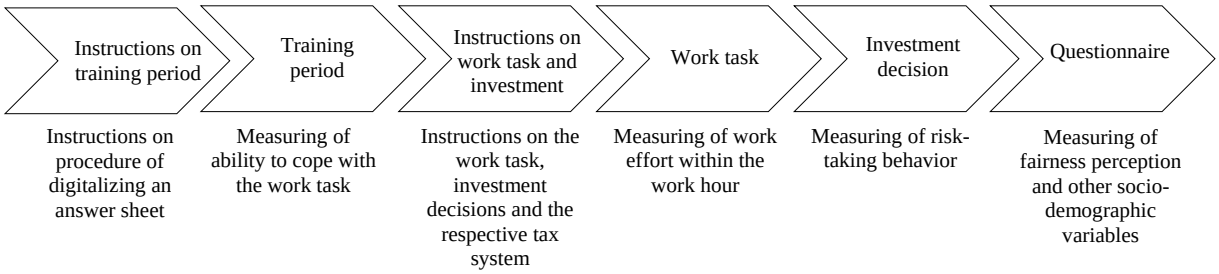
		Riskier Lottery			Less Risky Lottery		
Decision Period	State	1	2	3	1	2	3
	Probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
1	Rate of Return	60%	30%	0%	40%	30%	20%
2		70%	40%	10%	50%	40%	30%
3		50%	30%	10%	30%	30%	30%
4		40%	20%	0%	30%	20%	10%
5		60%	30%	0%	40%	30%	20%

The portfolio decisions occur in this fixed order. However, the presentation of the riskier lottery and less risky lottery varies. Thus, in the second and fifth decision periods, the riskier lottery is presented on the monitor's right-hand side, but in all other periods, it is presented on the left-hand side. The participants in the Immediate Taxation Treatment are informed that the return of their portfolio decision is tax free. In contrast, the participants in the Deferred Taxation Treatment are informed that a tax with a rate of 60% on the return and invested amount is withheld after their decisions.

After the investment phase, the participants are asked to answer a questionnaire that collects socio-demographic data and information on the perceived wage's fairness. A translated version of the questionnaire is presented in Section 7.9 (Appendix C). For payout purposes, only one portfolio decision is relevant. To determine the relevant return rate, the participant must throw the dice twice. The first throw decides on the decision period, and the second determines the state of environment. The participants are successively and separately paid out cash, whereby the payment is rounded up to the next ten cents. A timeline of the experimental procedure is illustrated in Figure 7.2.

[6] For a literature overview of the house money effect, see Clark (2002) as well as Weber and Zuchel (2005).

Figure 7.2: Timeline of Experimental Protocol



It is important to note that our immediate versus deferred taxation design does not consider the concept of the "time value of money" because this concept does not fit our experimental approach. The participants receive one net payment at the end of the experiment. Thus, there is no real time effect at all. For a similar approach, see Falsetta et al. (2013). Accordingly, we neither expect hyperbolic discounting nor uncertainty to influence our results. We argue that the time effect, which occurs in practice, would rather strengthen our result because the probability of different mental accounts would increase due to the delayed payments.

7.3.2 Sample

A total of 121 students (49 females and 72 males) participated in the experiment. The subjects were 23.4 years on average, and 46.3% studied either in the Faculty of Economics and Management or in the Faculty of Philosophy.[7] There are no significant differences in the individual characteristics between the treatments. The subjects earned € 19.78 on average in approximately 120 minutes (approximately € 9.89 per hour), with a range from € 9.80 to € 46.40.

7.4 Results

7.4.1 Work Effort

7.4.1.1 Bivariate Analyses

We start our analysis by examining the impact of the tax system on work effort (Hypothesis 1). The variable WORK EFFORT is measured by the number of answer sheets that the subjects correctly transfer to the computer. Besides, we have also utilized the number of answer sheets independent of their correctness as work effort measure. All results presented in this paper remain qualitatively unchanged. The descriptive statistics for WORK EFFORT are displayed separately for the immediate Taxation and Deferred Taxation Treatments in Table 7.3. On average, 19.3 correct sheets were digitized in the Immediate Taxation Treatment, whereas 18.6 correct sheets were transferred in the Deferred Taxation Treatment. This difference, however, is not significant (Mann-Whitney U, $p = 0.714$). Thus,

[7] All other subjects studied in the Faculty of Mathematics and Physics, the Faculty of Natural Sciences, the Faculty of Engineering, the Faculty of Electrical Engineering and Informatics, the Faculty of Architecture, or the Faculty of Law.

we cannot confirm the first hypothesis that deferred taxation affects work effort. So far, our results are in line with the neoclassical neutrality prediction. However, as noted previously, this result does not inevitably mean that subjects do not use mental accounting. Rather, subjects may use mental accounting, while perceived income, perceived substitution, and fairness effects offset one another.

Table 7.3: Overview of WORK EFFORT Categorized by Treatment

	WORK EFFORT	
Treatment	Immediate Taxation	Deferred Taxation
Mean	19.30	18.60
Median	19.00	19.00
Standard Deviation	5.66	4.51
Minimum	11	10
Maximum	33	29
No. of Subjects	61	60
Mann-Whitney U test	p = 0.714	

Note: The table presents key figures of the variable WORK EFFORT. It depicts the number of correctly digitized answer sheets within the work hour and is displayed for each taxation treatment separately. We use a two-tailed Mann-Whitney U test to examine whether there is a significant difference between the treatments.

To test the use of mental accounts, we collect data on perceptions of payment fairness in the post-experimental questionnaire (Hypothesis 2a). We asked the participants at the end of our experiment (after they decided on the work effort and risky investment). Thus, the participants were aware of the taxation system not only because of the given instructions but also because of their experiences during the actual experiment. Using a 10-point scale, subjects were asked to report the perceived fairness of the work task's payment (1 = not fair at all and 10 = totally fair).

Table 7.4 reveals a significant difference in the perceptions of payment fairness between the Immediate and Deferred Taxation Treatments. The mean level of perceived FAIRNESS is 5.02 in the Immediate Taxation Treatment and 6.52 in the Deferred Taxation Treatment. By using a two-tailed Mann-Whitney U test, we find that this difference is strongly significant at a 1% level. Thus, the participants who must pay taxes on their wages perceive the payment as less fair than those who pay taxes at a later point (i.e., when deciding on their investment portfolio). As argued above, this result demonstrates that the subjects use different mental accounts for their wage income and their investment income. If subjects did not use mental accounting, subjects in the Deferred Taxation Treatment would (in line with neoclassical theory) comprehend that their wage is taxed indirectly when taxes are levied on their invested income. Thus, they should not demonstrate a different fairness' perception of their wages. However, as subjects underlie mental accounting, the subjects in the Deferred Taxation Treatment seem to evaluate the fairness of their payment using the pre-tax payment (€ 2) by ignoring the deferred taxation of their wage income in the investment phase, whereas the subjects in the Immediate Taxation Treatment use the after-

tax payment (€ 0.80). Hence, we can confirm Hypothesis 2a, which states that mental accounting causes the perceived fairness of wages to be higher under deferred taxation of wages.

Table 7.4: The Impact of Deferred Taxation on Fairness Perceptions

	FAIRNESS	
Treatment	Immediate Taxation	Deferred Taxation
Mean	5.02	6.52
Median	5	7
Standard Deviation	2.19	2.41
Minimum	1	1
Maximum	10	10
No. of Subjects	61	60
Mann-Whitney U test	p = 0.0008	

Note: The table presents key figures of the perceptions of payment fairness. FAIRNESS displays the individual's self-reported satisfaction of the work task's payment measured on a 10-point scale where 1 = not fair at all and 10 = totally fair. We use a two-tailed Mann-Whitney U test to examine whether there is a significant difference between the treatments.

Next, we analyze how perceived fairness affects work effort (Hypothesis 2b) in order to investigate the effect of the tax system on work effort by examining the indirect effects via perceived fairness. Regarding the fairness perception, we use a median split to divide subjects into two groups. All subjects who report the perceived fairness of the work's payment to be below the median of the total sample (i.e., less than or equal to 5, measured on a 10-point scale where 1 = not fair at all and 10 = totally fair) are categorized as LOW FAIRNESS. All subjects who report at least the median level of perceived fairness (i.e., at least 6 points) are categorized as HIGH FAIRNESS. Using these two groups, we can analyze whether fairness perceptions significantly influence work effort. An overview of the work effort within fairness groups and the respective bivariate test is given in Table 7.5.

Table 7.5 depicts WORK EFFORT measured as the number of correctly transferred sheets and categorized according to perceived fairness. Whereas the LOW FAIRNESS group digitizes 18.14 sheets on average within the work hour, the HIGH FAIRNESS group digitizes 19.70 sheets on average. The difference is slightly significantly different from zero (p = 0.061). Hence, we can conclude that participants who perceive a payment as less fair tend to work less. Conversely, we can confirm Hypothesis 2b, which states that the greater the perceived fairness is, the greater the work effort is. The results regarding Hypotheses 2a and 2b point to an indirect effect of the timing of taxation (Hypothesis 3). Under deferred taxation, subjects perceive their payment to be fairer, and this increased fairness perception leads to greater work effort. In Section 7.4.1.2, we examine whether this indirect effect is significant.

Table 7.5: The Impact of Perceived Fairness on WORK EFFORT

	WORK EFFORT	
Fairness perception	LOW FAIRNESS	HIGH FAIRNESS
Mean	18.14	19.70
Median	17	20
Standard deviation	5.20	4.95
Minimum	11	10
Maximum	33	32
No. of subjects	58	63
Mann-Whitney U test	p = 0.061	

Note: WORK EFFORT depicts the number of correctly digitalized sheets within the work hour. It is analyzed for two binary categories of FAIRNESS. Subjects are allocated to the group LOW FAIRNESS if they report the perceived fairness to be less or equal than 5 and to the group HIGH FAIRNESS in all other cases. We use a two-tailed Mann-Whitney U test to examine whether there is a significant difference between the low and high fairness group.

7.4.1.2 Mediation Analysis

In the previous section, we demonstrate how the timing of taxation affects perceived fairness and show that stronger perceptions of fairness increase work effort (confirming Hypotheses 2a and 2b). However, in contrast to Hypothesis 1, we do not yet find a direct effect of the tax system on work effort. Using a structural equation model (SEM), we test whether the tax system indirectly affects work effort if perceived fairness operates as a mediator variable, apart from a possible direct effect of the tax system on WORK EFFORT. Assuming sequential ignorability (Imai et al., 2013), we estimate the following two equations:

$$WORK\ EFFORT \quad = \alpha_{01} + \alpha_{11} DEFERRED\ TAXATION + \alpha_{21} FAIRNESS + \varepsilon_{01} \tag{7.3}$$

$$FAIRNESS = \alpha_{02} + \alpha_{12} DEFERRED\ TAXATION + \varepsilon_{02} \tag{7.4}$$

The widely used traditional, but rather conceptual, mediation analysis is based on Baron and Kenny (1986). A formal test related to Alwin and Hauser (1975) calculates the mediated or indirect effect by multiplying both direct effects: the direct effect of the tax treatment on fairness perception (coefficient α_{12}) and the direct effect of fairness perception on work effort (coefficient α_{21}). The calculation of asymptotic standard errors of the indirect effect is based on the multivariate delta method (Bishop et al., 2007; Sobel, 1982, 1986). Statistical significance is derived through a comparison of the indirect effect divided by the asymptotic standard errors to a standard normal distribution (see MacKinnon et al., 2007 for a review).

To estimate the simple linear SEM in one analysis, we use Stata 14 based on maximum likelihood and the Stata command “estat teffects” to decompose into direct, indirect, and total effects, again based on

the delta method (Stata Corp., 2015 p. 157; Sobel, 1987). The results are identical to the traditional approach.

Figure 7.3: A Structural Equation Model Linking the Tax System Treatment, Perceived Fairness, and Work Effort

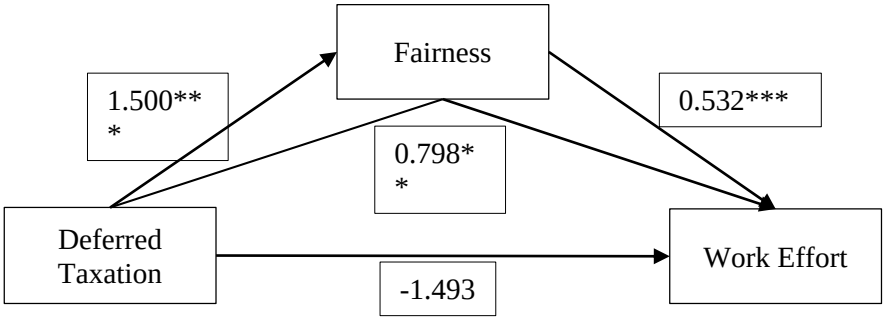


Table 7.6: Mediation Analysis

	FAIRNESS	WORK EFFORT		
		Direct	Indirect	Total
DEFERRED	1.500***	-1.493	0.798**	-0.695
TAXATION	(0.416)	(0.944)	(0.368)	(0.924)
FAIRNESS		0.532***		
		(0.196)		
CONSTANT	3.516***	18.120***		
	(0.655)	(1.574)		
No. of Observations	121	121	121	121
R-squared	0.116		0.062	
Wald Chi2	13.04***		7.95**	

Note: WORK EFFORT displays the number of correctly digitized answer sheets within the work hour. FAIRNESS displays the individual's self-reported satisfaction of the work task's payment measured on a 10-point scale where 1 = not fair at all and 10 = totally fair. DEFERRED TAXATION is a dummy variable equal to one if the observation belongs to the Deferred Taxation Treatment. The structural equation model includes equations (7.3) and (7.4). The indirect effect is calculated by multiplying both direct effects, i.e., the direct effect of the tax treatment on the fairness perception and the direct effect of the fairness perception on the work effort. The total effect is the sum of the direct and the indirect effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 7.3 and Table 7.6 display the results. As already presented in the previous section, the tax system has a strongly significant influence on perceived fairness ($p < 0.001$): as a result of mental accounting, subjects perceive the payment to be fairer under deferred taxation than under immediate taxation (Hypothesis 2a). FAIRNESS itself has a significant impact ($p = 0.007$) on WORK EFFORT, as greater fairness increases effort (Hypothesis 2b). We observe a significant indirect effect of the timing of taxation on WORK EFFORT, with FAIRNESS as a mediator ($p = 0.030$). Hence, we can confirm Hypothesis 3, which states that the Deferred Taxation Treatment indirectly increases work effort as a result of mental accounting because the perceived fairness of wages operates as a mediator variable.

We find neither a significant direct effect of DEFERRED TAXATION on WORK EFFORT ($p = 0.114$) nor a total effect as the sum of direct and indirect effects ($p = 0.452$). Thus, the SEM also fails to confirm Hypothesis 1.

Table 7.7: Mediation Analysis Including Socio-Demographic Variables and Ability

	FAIRNESS	WORK EFFORT		
		Direct	Indirect	Total
DEFERRED	1.700***	-0.595	0.672*	0.077
TAXATION	(0.423)	(0.926)	(0.359)	(0.886)
FAIRNESS		0.396**		
		(0.187)		
AGE	0.035	0.003	0.014	0.017
	(0.062)	(0.127)	(0.025)	(0.129)
MALE	0.300	-0.312	0.119	-0.194
	(0.477)	(0.870)	(0.176)	(0.868)
SOCIAL SCIENCE	0.101	-0.864	0.040	-0.824
	(0.005)	(0.852)	(0.165)	(0.868)
ABILITY	0.009**	0.041***	0.004	0.045***
	(0.005)	(0.010)	(0.003)	(0.010)
CONSTANT	4.211**	26.513***		
	(1.891)	(3.966)		
No. of Observations	121	121	121	121
R-squared	0.266		0.186	
Wald Chi2	27.63***		19.04***	

Note: WORK EFFORT displays the number of correctly digitized answer sheets within the work hour. FAIRNESS displays the individual's self-reported satisfaction of the work task's payment measured on a 10-point scale where 1 = not fair at all and 10 = totally fair. DEFERRED TAXATION is a dummy variable equal to one if the observation belongs to the Deferred Taxation Treatment. AGE displays participants' age measured in years. MALE is a dummy variable equal to one if the participant is male. SOCIAL SCIENCE displays whether a subject either studies at the faculty of economics or at the faculty of philosophy. ABILITY displays participants' required time to pass the practicing period. The structural equation model includes equations (7.3) and (7.4). The indirect effect is calculated by multiplying both direct effects, i.e., the direct effect of the tax treatment on the fairness perception and the direct effect of the fairness perception on the work effort. The total effect is the sum of the direct and the indirect effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To test whether these results are robust, we include socio-demographic control variables such as age (AGE), gender (MALE), faculty (SOCIAL SCIENCE), and ABILITY. ABILITY controls for the inherent and undistorted ability of the participant to perform the work task. ABILITY is measured as the time that participants need to finish the practicing period, multiplied by -1. Thus, a high value of ABILITY indicates a greater subject-specific ability to cope with the work task. Because the practicing period was conducted before the detailed treatment instructions were distributed (see Section 7.3.1), we ensure that the tax system does not affect the control variable ABILITY. We extend equations (7.3) and (7.4) and include AGE, MALE, SOCIAL SCIENCE and ABILITY as observed exogenous variables for our endogenous variables WORK EFFORT and FAIRNESS. Table 7.7 displays the results, which show that the previously reported effects remain the same. Additionally, we observe significant and plausible positive effects of ABILITY on WORK EFFORT (p-value < 0.001) as well as on perceived FAIRNESS (p-value = 0.041). None of the other control variables significantly affect perceived fairness or work effort.

7.4.2 Risk-Taking

7.4.2.1 Bivariate Analyses

In the following section, we examine whether the timing of taxation influences investment decisions (Hypothesis 4). As shown in Table 7.8, the results reveal no linear relation between the tax system and risk-taking. We observe a significant effect only if we consider decisions with high risk (i.e., decisions in which the percentage invested in the riskier asset exceeds a certain threshold, e.g., 2/3 or 3/4). In the Deferred Taxation System, the subjects make significantly fewer decisions with high risk (Mann-Whitney U, p = 0.002).

Table 7.8: The Impact of Deferred Taxation on Risk-Taking

Risk-taking Variable	Treatment	Mean	Median	SD	p-value
RISK-TAKING (CONTINUOUS)	Immediate	0.47	0.30	0.30	0.871
	Deferred	0.46	0.5	0.24	
RISK-TAKING (RISKY SHARE > 50%)	Immediate	0.35	0	0.48	0.372
	Deferred	0.31	0	0.46	
RISK-TAKING (RISKY SHARE > 66%)	Immediate	0.22	0	0.42	0.002
	Deferred	0.13	0	0.33	
RISK-TAKING (RISKY SHARE > 75%)	Immediate	0.19	0	0.39	0.002
	Deferred	0.10	0	0.30	

Note: The table presents key figures of the risk-taking variable under different conditions. The number of subjects in the Immediate (Deferred) Taxation Treatment amounts to 61 (60). RISK-TAKING (CONTINUOUS) denotes the average percentage invested in the riskier lottery, whereas RISK-TAKING (RISKY SHARE > 50% [60%] {75%}) takes the value 1 if the subject invests at least 50% [60%] {75%} in the riskier lottery and zero otherwise. The minimum (maximum) of all risk-taking variables amounts to zero (one). We use two-tailed Mann-Whitney U tests to examine whether there is a significant difference between the treatments.

7.4.2.2 Panel Analysis

The participants made five independent investment decisions in which they were asked to distribute their earned money into two lotteries with different levels of risk. To exploit the panel structure of our data and to meet the requirement of our left- and right-censored variable, we use the continuous RISK-TAKING variable[8] and run a random-effects tobit panel regression. We do not find a significant treatment effect (p = 0.633) independent of whether we control for socio-demographic variables such as age, gender, and faculty. Thus, we cannot confirm Hypothesis 4 when analyzing investment behavior with the continuous variable.[9]

Therefore, we again examine risk-taking behavior by analyzing decisions with high risk based on the variable HIGH RISK-TAKING, which indicated whether a subject invests at least 75% of the income

[8] By dividing the amount invested in the riskier lottery by the total earned amount, we calculate the share of income that was invested more riskily.

[9] However, if we consider only risk-taking decisions with positive contributions in the riskier lottery, we find a significant treatment effect on risk-taking; see section 7.4.2.4.

earned to the riskier lottery.[10] We use the variable HIGH RISK-TAKING as a dependent variable and run random-effects logit panel regressions.[11] The variable HIGH RISK-TAKING is a dummy variable equal to one if the subject invests at least 75% of her income into the riskier asset in period t (with t ranging from 1 to 5).

Table 7.9 displays two logit regressions run separately that include an increasing number of independent variables. Model 1 tests the influence of the tax system on the subjects' risk-taking behavior. Model 2 additionally controls for the socio-demographic variables AGE, MALE and SOCIAL SCIENCE. Additionally, we include the time needed for the investment decision for each period, DECISION $TIME_t$, as well as whether the riskier lottery was displayed on the monitor's left-hand side LEFT $SIDE_t$ as control variables.

The logit regressions in Table 7.9 demonstrate that the tax system significantly influences risk-taking behavior in both models ($p = 0.028$ and $p = 0.023$, respectively). Under deferred taxation, participants take less risk in investing (fewer decisions are made to invest at least 75% in the riskier lottery) than they do under the Immediate Taxation Treatment. Thus, when the tax is already levied directly on the income earned, people make more risky investments, whereas their investments are less risky when the income earned is tax free but the return and invested amount are taxed. This result supports Hypothesis 4. Model 2 reveals that, in line with prior research (Croson and Gneezy, 2009), we find a significant influence of MALE on risk-taking ($p = 0.010$). On average, male subjects make one more risky investment (i.e., investing at least 75% in the riskier lottery) than females in all five lotteries.

[10] We also created the HIGH RISK-TAKING variable with a 66% investment in the riskier lottery as a threshold. All results remain qualitatively unchanged but are not reported here.

[11] We find the same results when using a random-effects probit regression.

Table 7.9: Random-Effects Logit Regressions (Dependent Variable: HIGH RISK-TAKING$_t$)

	HIGH RISK-TAKING$_t$ (1)	HIGH RISK-TAKING$_t$ (2)
DEFERRED	-1.245**	-1.251**
TAXATION	(0.568)	(0.556)
AGE		-0.014
		(0.088)
MALE		1.490***
		(0.590)
SOCIAL SCIENCE		-0.750
		(0.546)
DECISION TIME$_t$		0.001
		(0.002)
LEFT SIDE$_t$		-0.262
		(0.331)
CONSTANT	-1.197	-1.276
	(0.831)	(2.322)
No. of observations	605	590
No. of subjects	121	118
Prob > chi2	0.028	0.050

Note: HIGH RISK-TAKING$_t$ is a dummy variable equal to one if the subject invests at least 75% of her income into the riskier asset in period t (with t = 1 to 5). DEFERRED TAXATION is a dummy variable equal to one if the observation belongs to the Deferred Taxation Treatment. AGE displays participants' age measured in years. MALE is a dummy variable equal to one if the participant is male. SOCIAL SCIENCE denotes whether a subject either studies at the faculty of economics or at the faculty of philosophy. DECISION TIME$_t$ displays the required time for the investment decision in period t. LEFT SIDE$_t$ is a dummy variable and takes the value 1 if the riskier lottery is presented on the monitors' left-hand side. In model 2, three subjects are not considered as we lack data on the decision time. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7.4.2.3 Diversification Heuristics

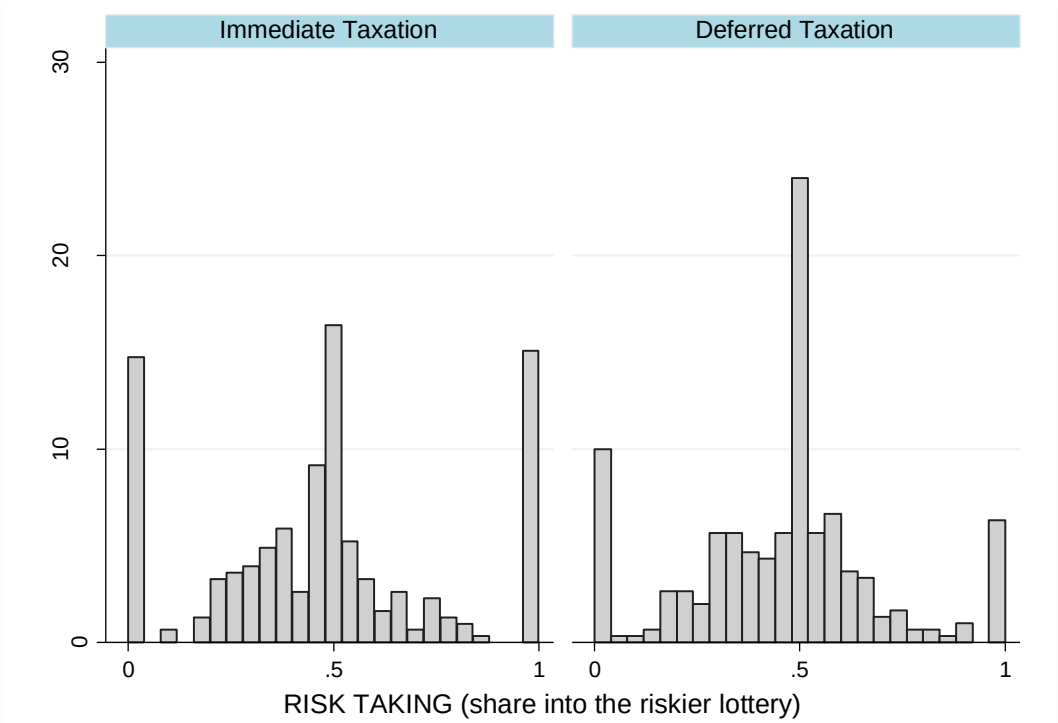
The result that deferred taxation leads to lower risk-taking corresponds to the behavioral argument that taxation leads to lower risk-taking because of the additional complexity induced by taxes (Ackermann et al., 2013). If subjects use mental accounting, they will notice that their returns and the invested amount are taxed in the Deferred Taxation Treatment. In the Immediate Taxation Treatment, however, the taxation occurs only in the work stage, such that subjects are not confronted with any further taxes in the investment stage. Thus, the investment decision is more complex under deferred taxation because taxes are levied there.

Higher complexity could also lead to an increase in the amount of time needed to make an investment decision. However, our results show that deferred taxation does not significantly extend the decision time.[12] This surprising result suggests that subjects might use decision heuristics to reduce the complexity in the environment. In line with the complexity argument, we expect that subjects more

[12] In the Immediate Taxation Treatment, the subjects need 53.7 seconds on average to make a single investment decision, whereas the subjects need 55.9 seconds on average in the Deferred Taxation Treatment. We perform a two-tailed Mann-Whitney U test to examine whether the decision time between the tax treatments differs significantly (p = 0.614).

frequently use simple heuristics in the Deferred Taxation Treatment. We present the distribution of RISK-TAKING in Figure 7.4.

Figure 7.4: Distribution of RISK-TAKING for each Treatment



As shown in Figure 7.4, risk-taking behavior differs between the two treatments. In the Immediate Taxation Treatment, the share of total investment decisions that is invested fully in the less risky lottery or fully in the riskier lottery (RISK-TAKING = 0 or RISK-TAKING = 1) is significantly higher than that in the Deferred Taxation Treatment ($p < 0.001$, chi2 test). Accordingly, the subjects in the Deferred Taxation Treatment diversify more between the two lotteries than the subjects in the Immediate Taxation Treatment. Note that in the used lottery setting it is not possible to minimize risk with diversification.[13] Nevertheless, the subjects seem to prefer diversification. This result is consistent with prior findings on the use of a diversification heuristic. The subjects display a desire for variety that leads to more diversity than they actually want (Read and Loewenstein, 1995; Hedesstrom et al., 2004). In particular, the number of decisions that distribute the earned income evenly between both lotteries is higher in the Deferred Taxation Treatment.

We observe that 12.5% of the investment decisions made in the Immediate Taxation Treatment are based on an even split of the income between the two assets. In contrast, 24% of the investment decisions in

[13] Assuming either a concave or convex expected utility function, risk seeking subjects should always invest 100% in the riskier lottery, whereas risk averse subjects should invest 100% in the less risky lottery. If a subject rather shows a risk utility function that is increasing in small variances but decreasing in larger variances, an interior solution may be optimal. (For a justification of Friedman-Savage utility functions even in perfect capital markets, see Hartley and Farrell, 2002). An interior solution can further be optimal if a subject demands a minimum rate of return above the worst outcome of the riskier lottery but, apart from that, seeks risk. However, these interior solutions do not inhibit diversification characteristics.

the Deferred Taxation Treatment are based on this even distribution ($p < 0.001$, chi2 test and Fisher's exact test). We use the aggregated application of the even split heuristic to test whether we observe statistically significant differences between the two tax system treatments. HEURISTIC represents the individual's heuristic behavior measured on a 6-point scale, where 0 = the subject never uses the even split heuristic and 5 = the subject uses this heuristic in all 5 periods. A two-tailed Mann-Whitney U test reveals a significant difference between the two tax system groups (p-value = 0.021). Moreover, we use the panel structure and run random-effects logit regressions with $HEURISTIC_t$ as an independent variable (a dummy variable equal to one if the subject uses the even split heuristic in period t, with t ranging from 1 to 5; p-value = 0.013). The higher tax complexity in the DEFERRED TAXATION TREATMENT appears to be accompanied by a significant increase in the use of the diversification heuristic, especially the even split heuristic (see Benartzi and Thaler, 2001 for the widespread use of the 1/n heuristic). We interpret the greater application of these special heuristics as a proxy for higher complexity and argue that this increased complexity leads to a lower propensity for risk-taking in the Deferred Taxation Treatment.

7.4.2.4 Panel Regressions with Regard to Diversification Heuristics

In the previous section, we demonstrated that subjects in the Immediate Taxation Treatment invest their total earned money in the less risky lottery significantly more often because they do not use the diversification heuristic as frequently as subjects in the Deferred Taxation Treatment. If we exclude all decisions in which the share invested in the riskier lottery is zero (RISK-TAKING = 0) and run the random-effects tobit panel regression again, we find that risk-taking differs significantly between the two treatments. These results are illustrated in Table 7.10. We find that risk-taking is significantly greater in the Immediate Taxation Treatment. Again, we additionally find that male subjects show a higher propensity for risk-taking than females do.

Table 7.10: Random-Effects Tobit Regressions (Dependent Variable: RISK-TAKING$_t$)

	RISK-TAKING$_t$ (1)	RISK-TAKING$_t$ (2)
DEFERRED	-0.055*	-0.067**
TAXATION	(0.033)	(0.033)
AGE		-0.005
		(0.005)
MALE		0.094***
		(0.033)
SOCIAL SCIENCE		-0.039
		(0.033)
DECISION TIME$_t$		0.000
		(0.000)
LEFT SIDE$_t$		-0.016
		(0.015)
CONSTANT	0.620***	0.727***
	(0.512)	(0.133)
No. of Observations	530	515
No. of Subjects	118	115
Prob > chi2	0.094	0.029

Note: RISK-TAKING$_t$ is the percentage invested in the riskier lottery in period t (with t = 1 to 5). Only observation where RISK-TAKING$_t$ >0 are included. DEFERRED TAXATION is a dummy variable equal to one if the observation belongs to the Deferred Taxation Treatment. AGE displays participants' age measured in years. MALE is a dummy variable equal to one if the participant is male. SOCIAL SCIENCE denotes whether a subject either studies at the faculty of economics or at the faculty of philosophy. DECISION TIME$_t$ displays the required time for the investment decision in period t. LEFT SIDE$_t$ is a dummy variable and takes the value 1 if the riskier lottery is presented on the monitor's left-hand side. In model 2, three subjects are not considered as we lack data on the decision time. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7.5 Robustness Checks

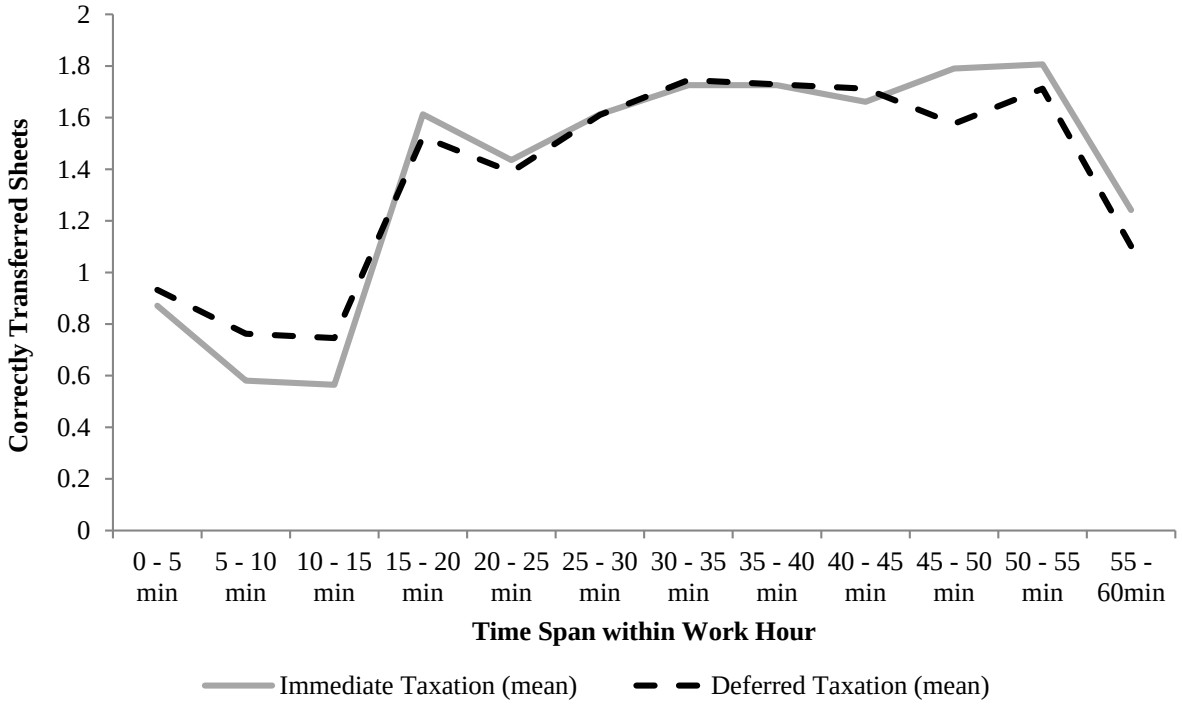
In this section, we subject our results to a series of robustness checks. In particular, we study the development of work effort over time, examine whether DEFERRED TAXATION indirectly affects RISK-TAKING (moderated by WORK EFFORT), and investigate whether our results are affected by subjects misunderstanding the instructions.

7.5.1 Work Effort over Time

Thus far, we have analyzed the impact of deferred taxation on work effort by examining the number of correctly transferred sheets within one work hour. However, the tax system might not only influence the output as total work effort but also trigger productivity effects over time. Thus, subjects might start working hard much earlier under one tax system, whereas they might work longer under the other system. Therefore, to control for those effects, we analyze productivity over time.

First, we cluster work time into time spans of five minutes, thus generating twelve points of time at which we measure the number of correctly transferred sheets. Second, we determine the number of sheets per time span. The results are illustrated in Figure 7.5 and separated by tax system. The findings reveal only minor differences in the number of digitized sheets between the treatments that are never significant (analyzed with a Mann-Whitney U test). Thus, we can conclude that deferred taxation does not directly influence productivity at a certain point within the work hour.

Figure 7.5: Number of Correctly Transferred Sheets over Time



7.5.2 Structural Equation Modelling including Risk-Taking

In the next step, we integrate the investment decision into the SEM to account for WORK EFFORT, which might also influence risk-taking beyond the timing of taxation. The existing literature reveals that risk-taking behavior is likely influenced by a house money effect. In this context, effort could negatively influence risk-taking (Arkes et al., 1994). The main results of the SEM are presented in Table 7.11. To integrate high risk-taking in the SEM, we generate a binary variable that takes the value 1 if the subject invests at least 75% of her income in the riskier lottery and 0 otherwise for all five decision periods. Subsequently, we sum these variables over all periods to create the variable HIGH RISK-TAKING. Depending on the single decision periods, HIGH RISK-TAKING can take values from 0 (i.e., a subject never invests at least 75% of her income in the riskier asset) to 5 (i.e., a subject invests at least 75% of her income in the riskier asset in all five decision periods). Note that we do not use ABILITY as control variable for HIGH RISK-TAKING, as the ability to cope with the work task correlates with work effort. We observe a significant negative effect of the tax system treatment on high risk-taking ($p = 0.014$). Hence, we again confirm Hypothesis 4, as subjects under deferred taxation make less risky investments because of the tax-induced complexity. Moreover, we find that work effort negatively influences risk-taking behavior (p-value = 0.073), which is consistent with prior findings: greater effort leads to an increase in risk aversion.[14] The results of the SEM's work effort analysis remain unchanged.

[14] Considering the context of tax evasion decisions, Kirchler et al. (2009) find that subjects evade more money if only low effort was required to earn it. Hence, as a tax evasion decision can be tantamount to a risk-taking decision, we should expect subjects who work harder to take less risk.

Table 7.11 Mediation Analysis Including Socio-Demographic Variables and Ability, Including Risk-Taking Behavior

	FAIRNESS	WORK EFFORT			HIGH RISK-TAKING
		Direct	Indirect	Total	
DEFERRED TAXATION	1.700***	-0.595	0.672*	0.077	-0.525**
	(0.423)	(0.926)	(0.359)	(0.886)	(0.214)
FAIRNESS		0.396**			
		(0.187)			
WORK EFFORT					-0.037*
					(0.021)
AGE	0.035	0.003	0.014	0.017	-0.009
	(0.062)	(0.127)	(0.025)	(0.129)	(0.032)
GENDER	0.300	-0.312	0.119	-0.194	0.679***
	(0.477)	(0.870)	(0.176)	(0.868)	(0.214)
SOCIAL SCIENCE	0.101	-0.864	0.040	-0.824	-0.303
	(0.005)	(0.852)	(0.165)	(0.868)	(0.986)
ABILITY	0.009**	0.041***	0.004	0.045***	
	(0.005)	(0.010)	(0.003)	(0.010)	
CONSTANT	4.211**	26.513***			2.154**
	(1.891)	(3.966			(0.986)
No. of Observations	121	121	121	121	121
R-squared	0.266		0.186		
Wald Chi2	27.63***		19.04***		18.10***

Note: WORK EFFORT displays the number of correctly digitized answer sheets within the work hour. FAIRNESS displays the individual's self-reported satisfaction of the work task's payment measured on a 10-point scale where 1 = not fair at all and 10 = totally fair. HIGH RISK-TAKING$_t$ is a dummy variable equal to one if the subject invests at least 75% of her income into the riskier asset in period t (with t = 1 to 5). DEFERRED TAXATION is a dummy variable equal to one if the observation belongs to the Deferred Taxation Treatment. AGE displays participants' age measured in years. MALE is a dummy variable equal to one if the participant is male. SOCIAL SCIENCE displays whether a subject either studies at the faculty of economics or at the faculty of philosophy. ABILITY displays participants' required time to pass the practicing period. The structural equation model includes equations (7.3) and (7.4). The indirect effect is calculated by multiplying both direct effects, i.e., the direct effect of the tax treatment on the fairness perception and the direct effect of the fairness perception on the work effort. The total effect is the sum of the direct and the indirect effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7.5.3 Instructions Check

In the post-experimental questionnaire, we ask the participants in the Deferred (Immediate) Taxation Treatment to state the amount of their wage after (before) taxation per sheet, and we ask those in the Deferred (Immediate) Taxation Treatment to report the tax rate with respect to the taxation of the invested amount and the return (wage). Eleven participants fail to answer both questions correctly. To conduct a robustness test, we run SEM regressions and random-effects logit regressions that include a further control variable named CHECK. This variable is a dummy variable equal to one if the subject answers both questions correctly and zero otherwise. Additionally, we exclude the participants who fail to answer the questions correctly and re-run all regression models. The results remain qualitatively unchanged. We observe a highly significant effect of the tax system treatment on perceived fairness, a significant direct effect of fairness on work effort and a significant indirect effect of the tax system treatment on work effort via fairness as a mediator. The direct effect of deferred taxation on work effort remains non-significant. Moreover, we find that deferred taxation leads to a significantly lower amount of risky investments (measured as 75% dummy variable and 66% dummy variable).

7.6 Conclusion

Using a real-effort laboratory experiment, we study whether the immediate or deferred taxation of compulsory pension contributions affects labor and investment decisions. In contrast to neoclassical predictions but in line with the assumption that individuals use different mental accounts for their work income and their investment income, we find that deferred taxation of wage income results in greater perceptions of fairness of the wage payment. This perception in turn increases participants' work effort. Thus, the timing of taxation (immediate versus deferred taxation) indirectly affects work effort. Furthermore, also in contrast to the neoclassical prediction, the results show that risk-taking is lower under deferred taxation.

Mental accounting seems to be the best explanation for our results. First, if neoclassical theory held we should neither expect any different perception of the payment's fairness nor different risk-taking behavior. Second, as we model our experiment in such a way that participants receive their payment in both treatments at the same time (right after the experiment), time-inconsistent preferences such as hyperbolic discounting cannot explain treatment differences. Moreover, due to the given certain tax rates in the experiment, we can exclude that uncertainty regarding future tax rates on pension income under deferred taxation confound our results. An important point, however, are potential learning effects. We designed our experiment as a one-shot game and thus refrained from analyzing learning effects. We think that this is an appropriate first step because subjects usually do not experience the taxation of their pension income until their retirement. Note that we use a simple and salient tax system and subjects are given all information needed to make their optimal decisions. However, one might argue that subjects may learn by observing the pension taxation of other subjects, e.g., their parents. Thus, it may be worth investigating the effects of such learning opportunities on the observed differences between deferred and immediate taxation in future research.

Our study contributes to the behavioral tax research that emphasizes the importance of incorporating psychological insights into the economic analysis of tax policy. Our study reveals that the assumption of neutrality regarding the timing of taxation does not hold. This finding has consequences for many tax policy issues. In the field of pension taxation, deferred and immediate taxation compete. Under immediate taxation (TEE), all income earned is taxed in the respective year, even if it is paid into a pension plan. However, the withdrawal is untaxed. In contrast, under deferred taxation (EET), contributions are deductible, but withdrawals are fully taxed. Although prior (neoclassical) research regards both pension tax systems as equivalent, our findings suggest that policy makers should consider that both pension tax systems might lead to different tax revenues because behavioral responses differ between the two systems as a result of mental accounting. While our study investigates only the effect on work effort and risk-taking, one might expect also differences concerning the effect on subjects' savings. If subjects use different mental accounts for their current and future income, they may not pay sufficient attention to the full taxation of their future income under a deferred taxation system. This

oversight could lead to lower savings (and thus lower after-tax pensions) in a deferred taxation system compared with an immediate taxation system, as observed by Beshears et al. (2017). Moreover, future research could examine the effects of mental accounting in other tax policy fields, such as dividend and capital gain taxation, where neoclassical predictions also use the neutrality concept in considering the timing of taxation.

7.7 Appendix A: Experimental Instructions

We divided the instructions into two parts. The first part is identical for both treatments and describes the procedure and training periods. The second part partly differs between the treatments. In the following, the instructions (originally written in German) are presented.

A1 Part 1 of the instructions

Hello, and welcome to our experiment!

By participating in this experiment, you have the possibility of earning money being provided by the University of Hanover. The aim of the experiment is to analyse economic decision making. The payment you receive at the end of the experiment depends on your effort as well as on chance. Please read the instructions carefully and attentively.

If you have further questions, please contact the experimenter.

1. Procedure

We would like to point out that you are not allowed to talk to the other participant or to leave your seat during the experiment.

You received a table tennis ball with an identification number to start the experiment. Please carefully keep the ball with you. You need the ball to identify yourself as soon as the compensation is paid. The identification number enables you to hide your true identity.

2. Training period

The experiment starts with a training period in which you get to know the design and functionality of the experiment. This period is designed to help you in the experiment.

Your task is to digitise the answers marked on the sheets in front of you into an entry form on the computer. The sheets contain the answers from a multiple choice exam. First, we ask you to enter the number of the sheet, which can be found at the top left corner of the page, into the field provided for it and press “Next.” Afterward, you will see the entry form for the sheet. It will be set up similar to the hard copy of the sheet in front of you. Please translate the marked answers for all 60 questions into the entry form on the computer. When you have finished translating the sheet, please press “Next.” Then, you will be given information on whether or not you correctly digitised on the sheet.

After every participant has finished the training period, you will be given a sign by the instructor. At that point, please open the envelope on your desk. Do not open the envelope before that.

A2 Part 2 of the instructions

1. Design of the experiment

[Immediate Taxation Treatment only: The experiment consists of two parts: You start with a one-hour working period in which you earn your wage. This wage is taxed immediately. Therefore, the experimenter will withhold the tax. The tax rate is 60% (for further information, please refer to point 3). In the second part of the experiment, you decide how you intend to invest your net wage (your wage after tax). The return on any investment (interest earned) is tax free.]

[Deferred Taxation Treatment only: The experiment consists of two parts: You start with a one-hour working period in which you earn your wage. In the second part of the experiment, you decide how you intend to invest your wage. The return on any investment and the invested amount must be taxed. The tax rate is 60% (for further information, please refer to point 3).]

As a last step, you will be asked to answer a short questionnaire, which is needed to evaluate the experimental results. We would like to emphasize that all answers will be kept anonymous. You will receive your payment after all participants have finished the experiment.

2. Working period

You can see a paper pile in front of you. These are answer sheets to a multiple choice exam that you have to digitise. This working task is exactly the same as during the previous training period.

Each answer sheet is numbered. You can find the sheet number at the top left corner. This sheet number is unique. Therefore, it is possible to identify the exam participant as well as the corresponding answer mask. As a consequence, it is extremely important that you type in the sheet number correctly!

First, we ask you to enter the number of the sheet, which can be found at the top left corner of the page, into the field provided for it and press “Next.” Afterward, you will see the entry form for the sheet. It will be set up similar to the hard copy of the sheet in front of you. Please translate the marked answers for all 60 questions into the entry form on the computer. When you have finished translating the sheet, please press “Next.” Then, you will be given information on whether or not you correctly typed in the sheet.

You get one hour to digitise as many sheets as you want or are able to do. During the whole period, a countdown will be projected onto the front wall of the room, telling you how much time is left. The experimenter will inform you once the hour has elapsed. We ask you to immediately stop digitizing the sheets at that moment, to look at the experimenter, and to press “Cancel” at his/her notice. Please stop digitizing sheets as soon as the working period is over!

If you want to take a break during the working period, you can use the internet, but you are not allowed to exit the room during the working period. You can access the internet by pressing the “Windows button” and clicking on the “Internet Explorer” symbol in the toolbar. Alternatively, you can access the

internet by pressing “tab” and “alt” at the same time. If you want to continue your work after your break, you have to press the “sheet” symbol in the bottom toolbar. Please do not press “Cancel” during the working period. Please wait until the experimenter tells you to do so. If you press “Cancel” prematurely, you are not able to continue the working task.

Please enter every sheet number just once. If you enter the same sheet number more than once, this sheet cannot be taken into consideration, and as a consequence, you will receive no compensation for this sheet. After the working period, you have the opportunity to make an investment decision.

3. Wage and taxation of wages

[Immediate Taxation Treatment only: You receive € 2.00 before taxation for each correctly digitized sheet; **60%** of this payment is withheld as tax, resulting in a payment of € 0.80 after taxation. After each digitized sheet, the computer informs you whether you entered the last sheet correctly and tells you the number of sheets you have entered correctly so far. At the beginning of the investment period, the number of sheets you entered, the number of sheets you entered correctly, the wage before taxation, the amount of taxes and the resulting wage after taxation are displayed.]

[Deferred Taxation Treatment only: You receive € 2.00 for each correctly digitized sheet. After each digitized sheet, the computer informs you whether you entered the last sheet correctly and tells you the number of sheets you entered correctly so far. In addition, you can see how much money you have earned so far. Your wage is not subject to any taxation.]

4. Investment period

[Immediate Taxation Treatment only: During this period of the experiment, you decide how to invest your wage after taxation.]

[Deferred Taxation Treatment only: During this period of the experiment, you decide how to invest your wage.]

Therefore, 2 different investment alternatives are displayed over 5 rounds. The return on the investment (interest) depends on the chosen investment alternative and the occurring state. Three possible states are given, with different impacts on your realized return. The probability for each state is equal to 1/3. For both investment alternatives, the same state occurs.

The investment alternatives are presented this way:

Investment A	State 1	State 2	State 3
Return	A%	B%	C%
Probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Investment B	State 1	State 2	State 3
Return	X%	Y%	Z%
Probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

[Immediate Taxation Treatment only: You are asked to split up the after-tax wage between the two investment alternatives. Therefore, please enter the amount you are willing to invest in investment alternative A and in investment alternative B (amounts in euros and cents). You can also invest your after-tax wage solely in one of the two alternatives. You have to invest all of your after-tax wage. The amount you have to invest remains constant over the 5 rounds. After the fifth round, one round is chosen for your payment. This decision will be made randomly by throwing a dice (for further information, please refer to point 6). Each payment you receive from the investment is subject to no taxation (for further information, please refer to point 5).

[Deferred Taxation Treatment only: You are asked to split up the wage between the two investment alternatives. Therefore, please enter the amount you are willing to invest in investment alternative A and in investment alternative B (amounts in euros and cents). You can also invest your wage solely in one of the two alternatives. You have to invest all of your wage. The amount you have to invest remains constant over the 5 rounds. After the fifth round, one round is chosen for your payment. This decision will be made randomly by throwing a dice (for further information, please refer to point 6). Each payment you receive from the investment is subject to taxation (for further information, please refer to point 5).

5. Taxation of investment

[Immediate Taxation Treatment only: The payment you receive from the investment is not subject to any taxation.]

[Deferred Taxation Treatment only: The payment you receive from the investment is subject to taxation. The tax rate is **60%**, meaning that the return on the investment and the invested amount is taxed at a rate of 60%. You will receive the after-tax payment (the tax is withheld by the experimenter).

6. Payment

Please answer the questionnaire as soon as you have finished your investment decisions. As mentioned above, we need this information to evaluate the results. All answers remain anonymous!

After each participant has completed the questionnaire, you will be asked to come to the front desk to receive your payment. Please use the dice to determine your return on investment.

The first roll of the dice determines the round of the investment decision. If 1 turns up, the investment decision made in the first round is taken into consideration, and if 2 turns up, the investment decision made in round two is taken into consideration, and so forth. If 6 turns up, you have to roll the dice again.

The second roll of the dice determines the occurring state. If 1 or 2 turns up, state 1 occurs. If 3 or 4 turns up, state 2 occurs. If 5 or 6 turns up, scenario 3 occurs.

If no further questions remain, please wait until the countdown for the working period to start. After that, please begin to work.

Thank you for participating!

7.8 Appendix B: Screenshots of the Experiment

Below are screenshots from the experiment. The experiment was conducted in German, hence, everything shown below was translated into English.

Figure 7.6: The Working Task (z-Tree Screenshot)

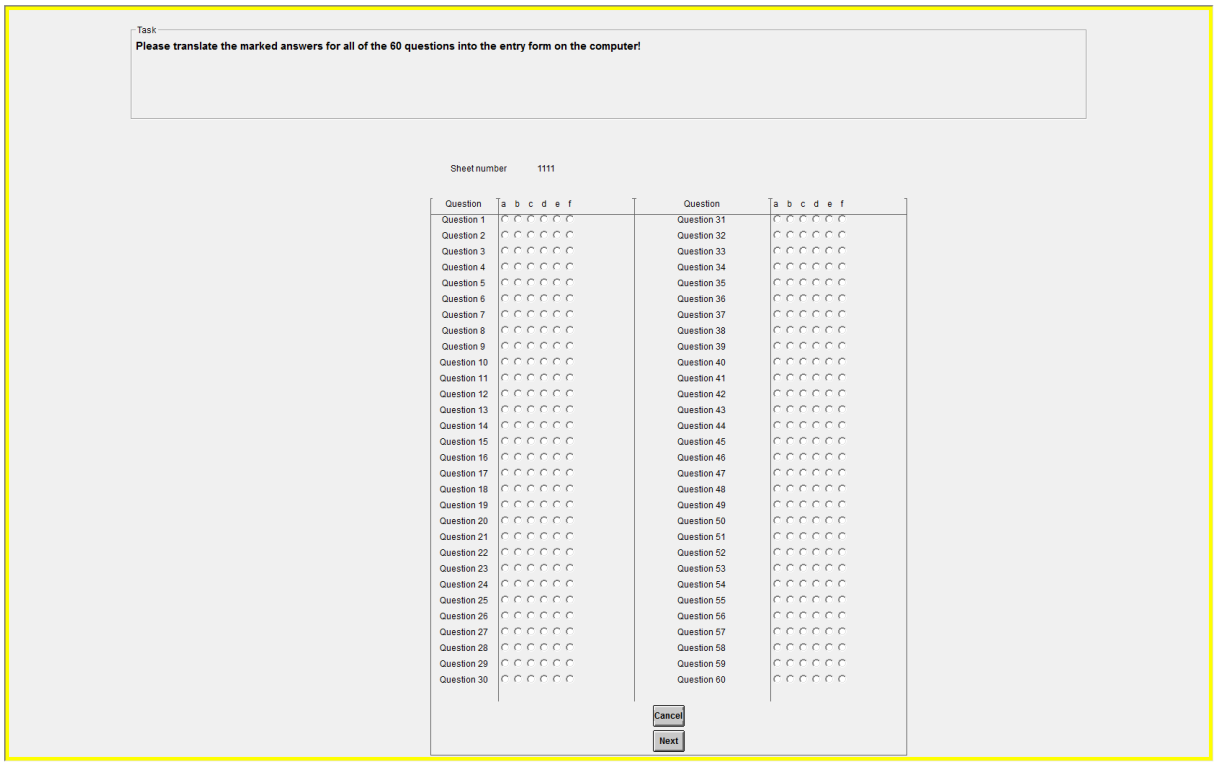


Figure 7.7: Information on Actual Work Progress (z-Tree Screenshot)

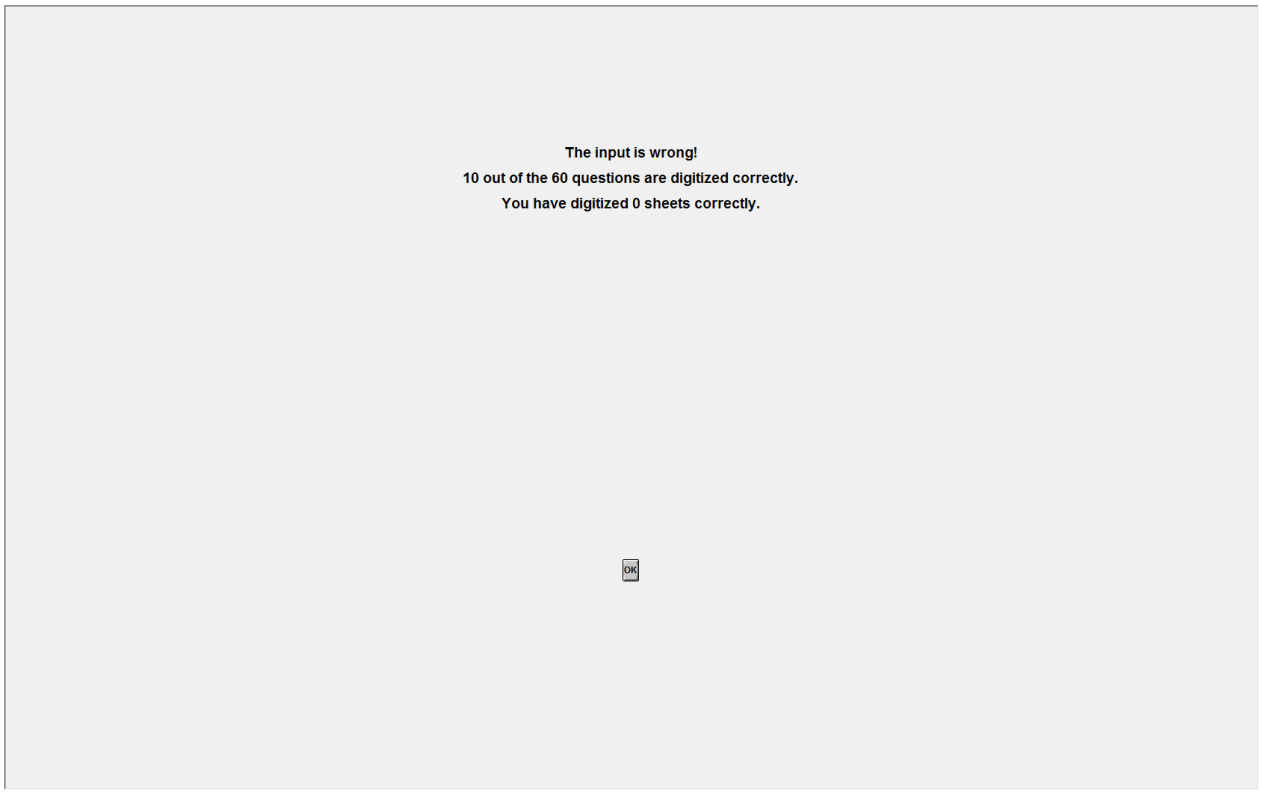


Figure 7.8: Entering a New Sheet Number and Information on Work Progress, Actual Wage before and after Tax, and the Tax Rate for the Immediate Taxation Treatment (z-Tree Screenshot)

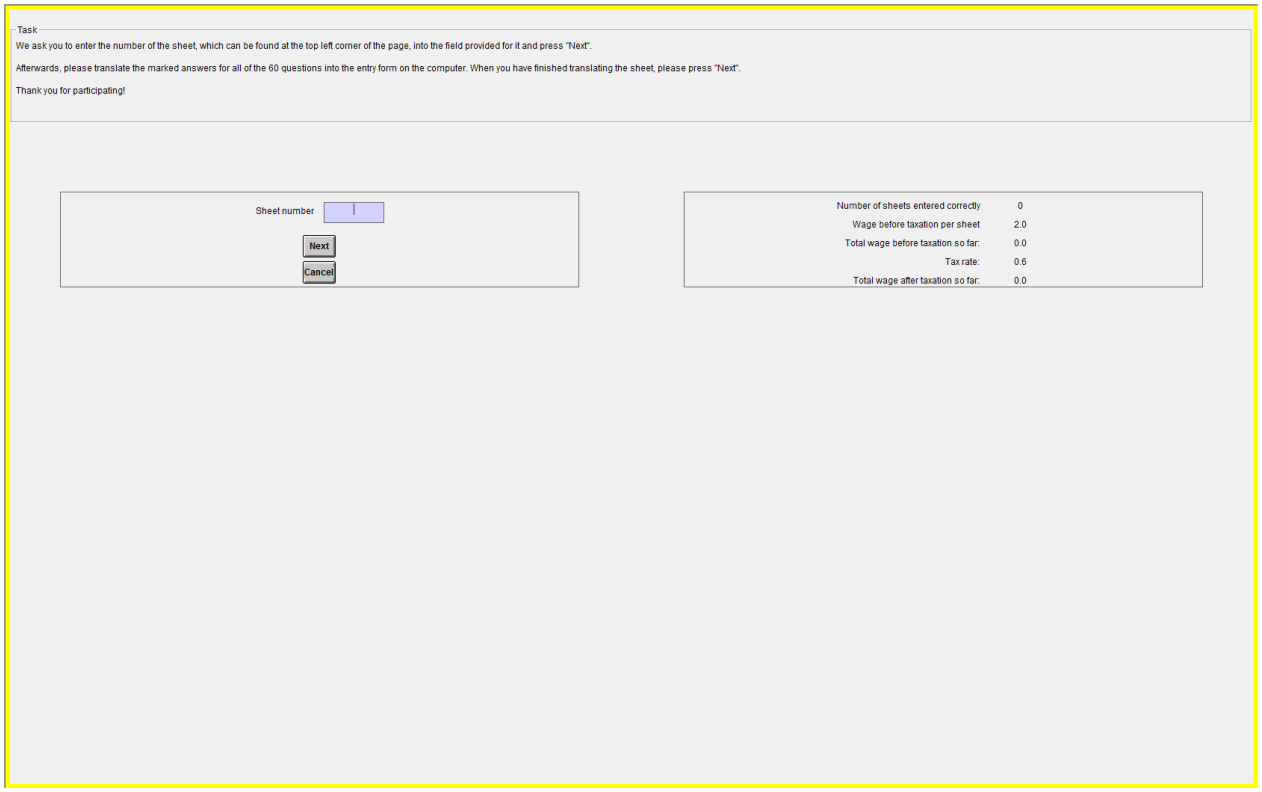


Figure 7.9: Entering a New Sheet Number and Information on Work Progress and Actual Wage for the Deferred Taxation Treatment (z-Tree Screenshot)

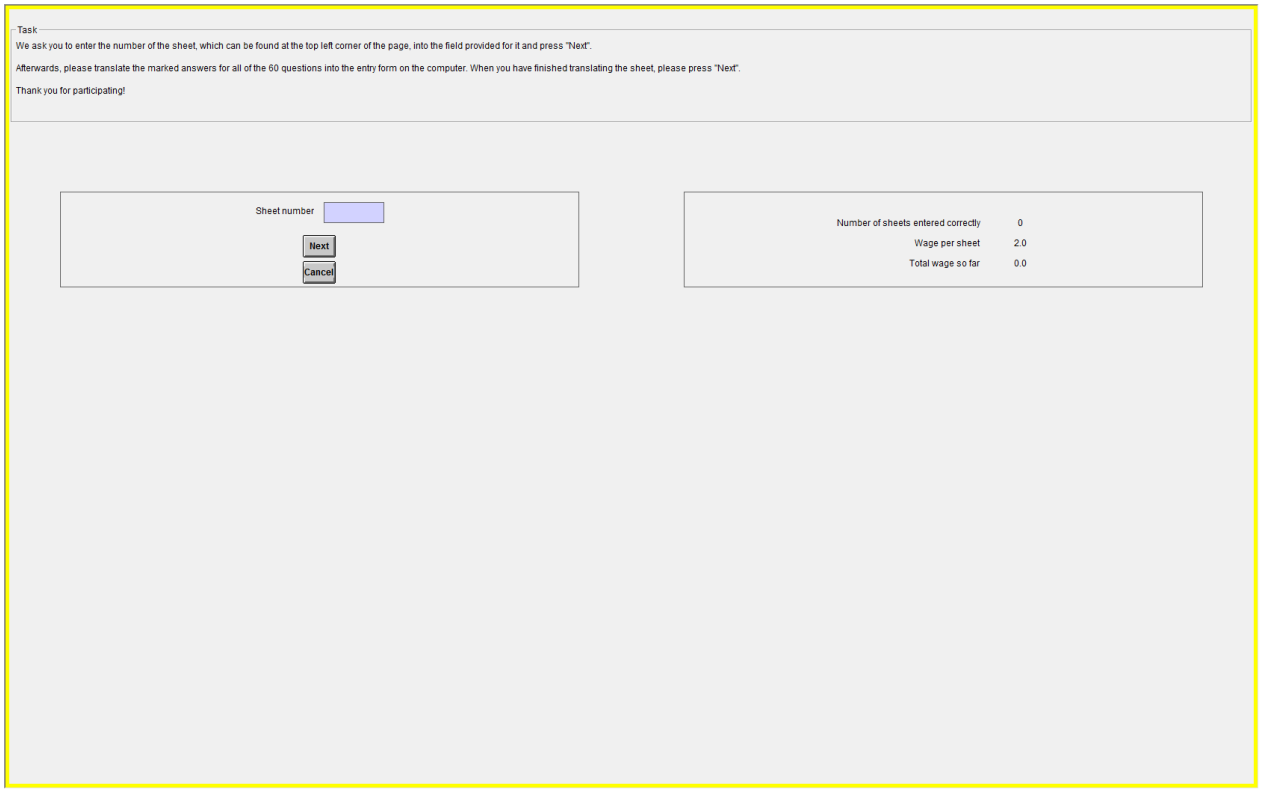


Figure 7.10: Investment Decision Stage Period 1 for the Immediate Taxation Treatment (z-Tree Screenshot)

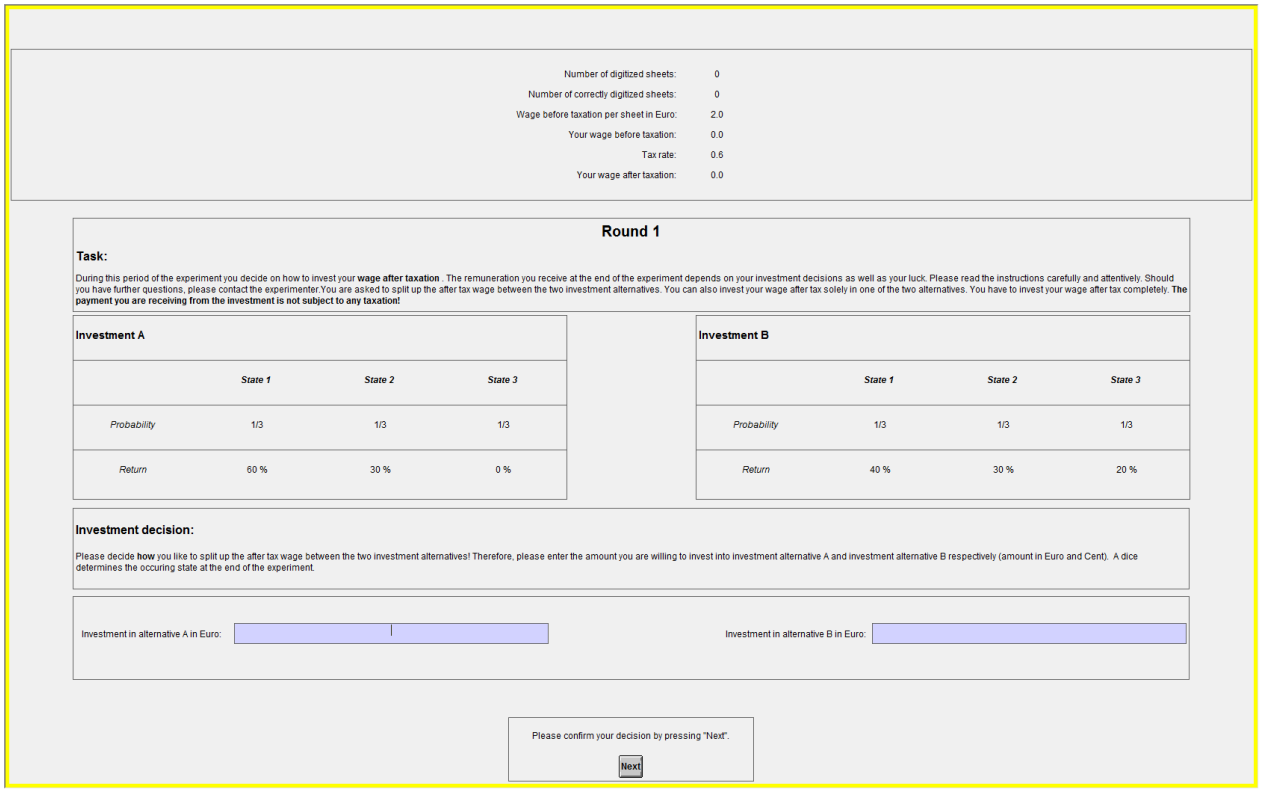
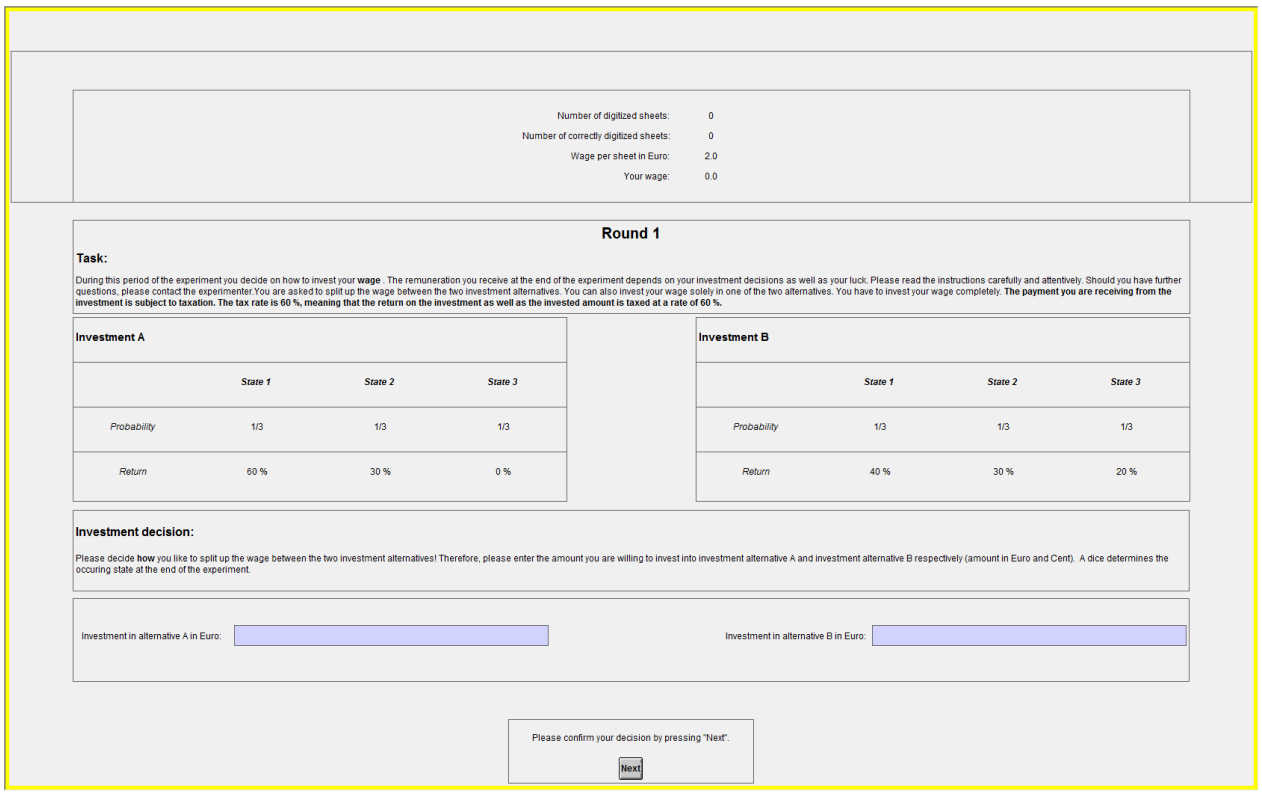


Figure 7.11: Investment Decision Stage Period 1 for the Deferred Taxation Treatment (z-Tree Screenshot)



7.9 Appendix C: Questionnaire

Below is the questionnaire from the experiment. The questionnaire was also conducted with z-tree and in German, hence, everything shown below was translated into English.

How did you perceive the working task during the experiment on a scale of 1 = very unpleasant to 10 = very pleasant?

Very unpleasant ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very pleasant

1 2 3 4 5 6 7 8 9 10

If you took a break during the working task, please tell us why!

- ☐ The task was unpleasant
- ☐ The wage was too low
- ☐ I had no more desire
- ☐ I did not take a break
- ☐ Others ☐

How important was the wage before taxation by determining the work effort on a scale of 1 = very unimportant to 10 = very important?

Very unimportant ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very important

1 2 3 4 5 6 7 8 9 10

Immediate *Taxation Treatment only:* Your wage was subject to taxation. What was the tax rate in percent?

☐ %

Deferred Taxation Treatment only: The return on the investment and the invested amount were subject to taxation. What was the tax rate in percent?

☐ %

How important was the taxation by determining the work effort on a scale of 1 = very unimportant to 10 = very important?

Very unimportant ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very important

1 2 3 4 5 6 7 8 9 10

Immediate *Taxation Treatment only:* Please tell us the amount of the wage after taxation per sheet in Euro.

☐ Euro

Deferred Taxation Treatment only: Please tell us the amount of the wage before taxation per sheet in Euro.

☐ Euro

Immediate *Taxation Treatment only:* How important was the wage after taxation by determining the work effort on a scale of 1 = very unimportant to 10 = very important?

Very unimportant ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very important

1 2 3 4 5 6 7 8 9 10

Deferred Taxation Treatment only: Did you recognize that the wage was effectively subject to taxation because of the taxation of the return on investment and the taxation of the invested amount?

☐ Yes ☐ No ☐ I don' t know

How did you perceive the fairness of the working task wage on a scale of 1 = not fair at all to 10 = totally fair?

Not fair at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Totally fair

1 2 3 4 5 6 7 8 9 10

How would you rate your tax law knowledge on a scale of 1 = no knowledge to 10 = exceptionally knowledge?

No knowledge ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Exceptionally knowledge

1 2 3 4 5 6 7 8 9 10

How would you rate your investment knowledge (e.g. investing in securities, bonds, fixed deposit accounts) on a scale of 1 = no knowledge to 10 = exceptionally knowledge?

No knowledge ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Exceptionally knowledge

1 2 3 4 5 6 7 8 9 10

Was the mean return the same for the two investments alternatives in each round?

☐ Yes ☐ No ☐ I don' t know

How old are you? ☐ (years)

Are you female or male?

☐ Female ☐ Male

Which faculty are you enrolled for?

- ☐ Architecture and landscape
- ☐ Construction engineering and geodesy
- ☐ Electrical engineering and computer science
- ☐ Law
- ☐ Mechanical engineering
- ☐ Mathematics and physics
- ☐ Natural sciences
- ☐ Philosophy, business and economics (social science)
- ☐ Others ____
- ☐ I am not a student

What qualification are you aiming at right now?

- ☐ Bachelor
- ☐ Master
- ☐ Diploma
- ☐ Magister [comparable to Master of Arts]
- ☐ 1st state examination
- ☐ 2nd state examination
- ☐ Doctoral degree
- ☐ Other ____

Which academic semester are you in? ____

What's your marital status?

- ☐ Marriage/registered partnership
- ☐ Unmarried
- ☐ Divorced/widowed

Do you have children?

- ☐ Yes
- ☐ No

What is your monthly disposable income (after rent; approximately)?

- ☐ < € 500
- ☐ € 501 - € 1.000
- ☐ € 1.001 - € 1.500
- ☐ € 1.501 - € 2.000
- ☐ > € 2.001

Chapter 8

Concluding Remarks

This thesis experimentally examines two major areas of interests within behavioral taxation for which empirical data hardly exists. Thereby, the first four papers analyze tax evasion behavior, and test tools which are considered to combat this crime. While the first two papers use group experiments to examine, inter alia, the influence of other taxpayers on personal tax evasion behavior, papers three and four incorporate individual choices.

Chapter two studies whether different forms of tax privacy impact tax compliance. Subjects are allocated into groups of five and have to make individual choices on their tax declaration which is paid into a public good and re-allocated among all group members. The experiment lasts fifteen periods. In the No Information Treatment subjects are not informed about the contribution of the other group members, whereas in the Partial Information Treatment they are informed about the other group members' contributions. However, this information is given anonymously so that they have no information on the actual subject behind the group members. In contrast, this information is given in the Full Information Treatment in which participants are additionally provided with photos of each member which are linked to the respective contribution. We find a shame effect which reduces tax evasion in the Full Information Treatment (compared to the Partial Information Treatment) in the short-run. Thus, providing detailed information on one's own tax compliance seems to hamper subjects from evading taxes. However, this effect diminishes over time which can be explained with the contagion effect. Comparing tax declarations in the Partial Information Treatment with the No Information Treatment we find, that tax evasion is significantly higher in the former treatment. This can be explained by a crowding out effect of formerly honest taxpayers if they are made aware of the dishonesty of others. As the shame effect is not strong enough to override the contagion effect in the long run, tax compliance is not sustainably increased by providing full information on others' tax compliance. In contrast, providing information anonymously even decreases tax compliance below the level than if no information was provided at all.

The third chapter examines tax compliance behavior of groups compared to individuals and whether a group allocation influences individual compliance afterwards. The experiment consists of nine periods (3x3 periods) and asks the subjects to declare a fictive company's income. While subjects in the Treatment I-I-I make all decisions individually, subjects in the Treatment G-G-G are allocated to a group

of three, discuss their decisions within the group and make the decision on their own afterwards. In the Treatment I-G-I subjects decide in the first three and in the last three periods individually but are allocated to a group in periods four to six. We find that tax compliance is significantly lower in the group settings than in the individual settings. This finding holds for both, the between subject comparison (Treatment I-I-I vs. G-G-G) and the within subject comparisons in Treatment I-G-I. However, analyzing Treatment I-G-I we find that group interaction induces a negative spillover effect on subsequent individual compliance so that tax compliance is significantly lower in the last three periods than in the first three. Thus, the contagion effect again decreases tax compliance as it causes a crowding out of formerly intrinsically motivated subjects. Interestingly, this behavioral adjustment is gender-dependent. Whereas male subjects do not adapt their already low compliance within or after the group allocation, females are more affected by the group allocation and reduce their tax compliance after the argumentation. In the subsequent individual choices they either stick to this low compliance level or increase compliance again. Over all group conversations risk was the predominant argument for encouraging or discouraging compliance behavior.

The fourth chapter analyzes tax evasion behavior in two possible channels: underreporting positive income or overdeducting negative income. Thereby, subjects are allocated to different individual choice experiments which, inter alia, differ in the occurrence of an offsetting-mechanism of both channels, and the salience level of this mechanism. Subjects robustly evade more tax by underreporting positive income than by overdeducting negative income. Although the effect size reduces due to the offsetting mechanism, the different tax compliance levels in both channels remain significant. This difference only vanishes if the offsetting mechanism is made very saliently, so that a conjoint decision on both tax evasion channels is almost unavoidable. Summarizing, the experiment provides evidence that first, tax compliance behavior is asymmetrically in case of positive and negative income reporting, second, offsetting of both income types reduces this effect and third, the degree of offsetting salience is decisive for the persistence of the effect.

Chapter five presents the last study on tax evasion and analyzes whether the self-disclosure opportunity affects tax compliance behavior. Objectors of a self-disclosure opportunity argue that it decreases tax compliance as it reduces tax morale and is used strategically in the tax evasion decision. In an individual choice framework, subjects decide on their tax declaration in the first five periods with certain audit probabilities and in the following seven periods given uncertain audit probabilities. Subjects are either allocated to the No Self-Disclosure Treatment or to one of two Self-Disclosure treatments which differ in the imposition of a possible penalty payment on the disclosed taxes. Whereas in the Self-Disclosure without Penalty Treatment no fine is levied, subjects have to pay a small fine in the Self-Disclosure with Penalty Treatment. The analysis of the first five periods reveals no different tax compliance behavior in all three treatments. Thus, the self-disclosure opportunity does not lead to an erosion of tax morale which would have caused a crowding out of formerly intrinsic motivated and honest subjects. The analysis of

periods with uncertain audit probabilities shows that subjects indeed integrate the self-disclosure option into their tax evasion decision, thus reducing tax compliance. However, small penalty payments that accompany the self-disclosure opportunity are sufficient to limit this tax evasion decrease. Considering that the self-disclosure option can be a useful tool to offer subjects a way back to tax honesty and that it can generate revenues that governments would not have been able to generate within their standard audit capacities otherwise, the self-disclosure opportunity that includes small penalty payments provides an instrument to increase tax revenues without bearing a considerable fiscal risk.

Summarizing, the allocation to a group in which subjects are aware of the tax compliance behavior of the others – either through direct discussion or through information on the others' tax compliance after the decisions – leads to a contagion effect as formerly honest taxpayers are made aware of the tax evasion behavior of others which leads to an erosion of tax morale and a crowding out of these formerly honest subjects. Although the provision of full information on the group members by revealing their identity induces a shame effect which works against the contagion effect, it is not sufficient to foster the shame effect in the long run. In contrast to both group experiments, we do not find a crowding out of formerly intrinsic motivated tax payers due to the introduction of a self-disclosure opportunity in an individual choice context. However, the triggers of the crowding out effect are different comparing groups' and individual choice experiments. While the contagion effect causes a crowding out due to the observation of others' dishonest behavior, we do not find a crowding out effect if a self-disclosure option is introduced which might have reduced tax morale if subjects perceive such an introduction as unjust or legalization of the tax evasion by the tax authority (e.g., Feld and Frey, 2002, 2007; Vihanto, 2003).

Furthermore, this thesis presents two studies on mechanisms to increase tax revenues: the abolition of tax privacy and the introduction of a self-disclosure opportunity. Taking into consideration that the positive shame effect diminishes over time and is overridden by the negative contagion effect, tax authorities should be cautious when considering the reduction of tax privacy rules. Regarding the self-disclosure opportunity, the study reveals that the linkage of rather small penalty payments on this opportunity is sufficient to limit the decrease of tax compliance.

Therefore, the thesis at hand gives recommendation to tax authorities on specific tax evasion drivers, as well as on limitations or specific considerations about tools to enhance tax compliance. However, these results must be considered against the background that they are based on laboratory measurements. Although the experiments are designed to investigate theoretical predictions while mirroring reality in the best possible way, it is hard to establish different facets of social norms in the lab. As tax compliance is only one such social norm which are likely interdependent, these interactions can only vaguely been considered or tested. However, these results give reliable information on behavioral aspects that have to be considered when discussing tax compliance.

The last two papers within this thesis use a real-effort task to generate observable data on the work effort in order to analyze the impact of different tax systems on various productivity parameters. In both

experiments subjects are asked to digitalize answers of a multiple choice exam. For each correctly transferred exam they earn money. In Chapter six, this money underlies different tax tariff systems that vary regarding their level of complexity. While the easiest tax system clearly states the effective tax rate, the most complex tax system combines four different tax rates with each other by means of different calculation options and interactions. The results show that tax rate complexity leads to an overestimation of the actual tax rate that decreases labor supply. Thus, in order to prevent taxpayers from reducing their work offer due to a misperception of the actual tax burden which directly impacts economic growth, tax authorities are advised to reduce complexity within the tax systems.

Chapter seven, in turn, analyzes two different systems of pension taxation. Thereby, subjects first complete the working task described above and finally invest their money into lotteries with different returns which reflects the investment into a random pension fund. While the wage is taxed right after the working task and only the net wage can be invested in the immediate tax system (the returns of the investment are tax-free), the wage is tax-exempt in the deferred tax systems but the returns of the lotteries as well as the invested amount is taxed in the very end. As both tax systems yield the same after-tax wealth, working and savings behavior should be unaffected by the timing of taxation according to neoclassical theory. However, subjects perceive their (untaxed) wage as significantly more fair in the deferred taxation treatment which indirectly increases their work effort. Additionally, subjects invest more riskily under this taxation system.

Concluding, the tax system's design regarding its complexity and the timing of taxation influences work effort. Thus, tax authorities can directly enhance labor supply by simplifying the tax rate's complexity and by delaying the timing of pension taxation to the future payoff period. Both measures would thereby have positive effects on the economy and increase the amount that can be spent by the individuals.

References

Abeler, J., and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3), 1-28.

Ackermann, H., Fochmann, M., and Mihm, B. (2013). Biased effects of taxes and subsidies on portfolio choices. *Economics Letters*, 120(1), 23-26.

Ackermann, H., Fochmann, M., and Wolf, N. (2016). The Effect of Straight-Line and Accelerated Depreciation Rules on Risky Investment Decisions—An Experimental Study. *International Journal of Financial Studies*, 4(4), 19-44.

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, 97(4), 543-569.

Akerlof, G. A. (1984). Gift Exchange and Efficiency-Wage Theory: Four Views. *American Economic Review*, 74(2), 79-83.

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

Alm, J. (1991). A perspective on the experimental analysis of taxpayer reporting. *The Accounting Review*, 577-593.

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

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., and Beck, W. (1990). Tax amnesties and tax revenues. *Public Finance Quarterly*, 18(4), 433-453.

Alm, J., and Beck, W. (1991). Wiping the slate clean: Individual response to state tax amnesties. *Southern Economic Journal*, 1043-1053.

Alm, J., and Beck, W. (1993). Tax amnesties and compliance in the long run: A time series analysis. *National Tax Journal*, 53-60.

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

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

Alm, J., McClelland, G.H., and Schulze, W.D. (1992). Why do people pay taxes? *Journal of Public Economics*, 48, 21-38.

Alm, J., McKee, M., and Beck, W. (1990). Amazing grace: Tax amnesties and compliance. *National tax journal*, 23-37.

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

Alwin, D. F., and Hauser, R. M. (1975). The Decomposition of Effects in Path Analysis. *American Sociological Review*, 40(1), 37-47.

Andersson, J., Schroyen, F., and Torsvik, G. (2019). The impact of international tax information exchange agreements on the use of tax amnesty: evidence from Norway. *NHH Dept. of Economics Discussion Paper*, (16).

Andreoni, J. (1991). The desirability of a permanent tax amnesty. *Journal of Public Economics*, 45(2), 143-159.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, 36(2), 818-860.

Andreoni, J., and Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, 88 (7-8), 1605-1623.

Ariely, D. (2012). The Honest Truth About Dishonesty: How We Lie to Everyone—Especially Ourselves, New York: Harper.

Arkes, H. R., Joyner, C. A., Pezzo, M. V., Nash, J. G., Siegel-Jacobs, K., and Stone, E. (1994). The Psychology of Windfall Gains. *Organizational Behavior and Human Decision Processes*, 59(3), 331-347.

Ashton, R. H., and Kramer, S. S. (1980). Students as Surrogates in Behavioral Accounting Research: Some Evidence. *Journal of Accounting Research*, 18, 1-15.

Atkinson, A. A., and Stiglitz, J. E. (1980). Lectures on Public Economics. London: McGraw-Hill.

Azar, O. H., Shira, Y., and Bar-Eli, M. (2013). Do customers return excessive change in a restaurant? A field experiment on dishonesty. *Journal of Economic Behavior and Organization*, 93, 219-226.

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.

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

Baldry, J.C. (1986). Tax evasion is not a gamble: A report on two experiments. *Economics Letters*, 22, 333-335.

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.

Barrera, D., and Simpson, B. (2012). Much ado about deception: Consequences of deceiving research participants in the social sciences. *Sociological Methods & Research*, 41(3), 383-413.

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.

Battigalli, P., and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97, 170-176.

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

Battiston, P., Duncan, D., Gamba, S., and Santoro, A. (2020). Audit Publicity and Tax Compliance: A Natural Experiment. *The Scandinavian Journal of Economics*, 122(1), 81-108.

Beck, P. J., and Jung, W. O. (1989). Taxpayers' reporting decisions and auditing under information asymmetry. *Accounting Review*, 468-487.

Becker, G. S. (1962). Irrational behavior and economic theory. *Journal of political economy*, 70(1), 1-13.

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

Belot, M., and Schröder, M. (2015). The spillover effects of monitoring: A field experiment. *Management Science*, 62(1), 37-45.

Bem, D. J. (1972). Self-perception theory. *Advances in experimental social psychology*, 6(1), 1-62.

Bénabou, R., Falk, A., and Tirole, J. (2019). Narratives, Imperatives, and Moral Persuasion. *Working Paper*.

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

Benartzi, S., and Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1) 79-98.

Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences. *Journal of the European Economic Association*, 11, 1231-1255.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2017). Does front-loading taxation increase savings? Evidence from Roth 401 (k) introductions. *Journal of public economics*, 151, 84-95.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (2007). Discrete Multivariate Analysis: Theory and Practice. New York: Springer Science & Business Media..

Blaufus, K., Bob, J., Hundsdoerfer, J., Kiesewetter, D., and Weimann, J. (2013). Decision heuristics and tax perception – An analysis of a tax-cut-cum-base-broadening policy. *Journal of Economic Psychology*, 35, 1-16.

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.

Blaufus, K., Hundsdoerfer, J., Jacob, M., and Sünwoldt, M. (2016). Does legality matter? The case of tax avoidance and evasion. *Journal of Economic Behavior & Organization*, 127, 182-206.

Blaufus, K., and Möhlmann, A. (2014). Security Returns and Tax Aversion Bias: Behavioral Responses to Tax Labels. *Journal of Behavioral Finance*, 15(1), 56-69.

Blaufus, K., and Ortlieb, R. (2009). Is Simple Better? A Conjoint Analysis of the Effects of Tax Complexity on Employee Preferences Concerning Company Pension Plans. *Schmalenbach Business Review*, 61, 60-83.

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.

Blumkin, T., Ruffle, B. J., and Ganun, Y. (2012). Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods. *European Economic Review*, 56(6), 1200-1219.

Bø, E. E., Slemrod, J., and Thoresen, T. O. (2015). Taxes on the internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy*, 7(1), 36-62.

Bobek, D. D., Hageman, A. M., and Kelliher, C. F. (2013). Analyzing the role of social norms in tax compliance behavior. *Journal of Business Ethics*, 115(3), 451-468.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71, 117-120.

Bolle, F., and Otto, P. E. (2010). A price is a signal: on intrinsic motivation, crowding-out, and crowding-in. *Kyklos*, 63, 9-22.

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

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics* 127, 1243-1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy* 121, 803-843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2015). Salience Theory of Judicial Decisions. *The Journal of Legal Studies* 44, 7-33.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2016). Competition for Attention. *Review of Economic Studies* 83, 481-513.

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.

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

Boylan, S. J., and Frischmann, P. J. (2006). Experimental Evidence on the Role of Tax Complexity in Investment Decisions. *Journal of the American Taxation Association*, 28, 69-88.

Break, G. F. (1957). Income Taxes and Incentives to Work: An Empirical Study. *The American Economic Review*, 47 (5), 529-549.

Calvet, L. E., and Sodini, P. (2014). Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios. *Journal of Finance*, 69(2), 867-906.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, 2(1), 61-104.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2012). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 238-249.

Cameron, J., Miller, D., and Monin, B. (2008). Deservingness and unethical behavior in loss and gain frames. *Working paper*, University of California, Berkeley.

Chambers, V., and Spencer, M. (2008). Does changing the timing of a yearly individual tax refund change the amount spent vs. saved? *Journal of Economic Psychology*, 29(6), 856-862.

Chang, O. H., and Schultz, J. J. (1990). The Income Tax Withholding Phenomenon: Evidence From TCMP Data. *Journal of the American Taxation Association*, 12, 88-93.

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.

Charness, G., and Kuhn, P. (2011). Lab Labor: What Can Labor Economists Learn from the Lab? *Handbook of Labor Economics*, 4a, 229-330.

Charness, G., and Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, 26, 157-176.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *The American Economic Review* 99, 1145-1177.

Chetty, R., and Saez, E. (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics*, 5 (1), 1-31.

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.

Chiappori, P., and Paiella, M. (2011). Relative Risk Aversion Is Constant: Evidence from Panel Data. *Journal of the European Economic Association*, 9(6), 1021-1052.

Christian, C. W., Gupta, S., and Young, J. C. (2002). Evidence on subsequent filing from the state of Michigan's income tax amnesty. *National Tax Journal*, 703-721.

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

Clark, J. (2002). House money effects in public good experiments. *Experimental Economics*, 5(3), 223-231.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., and Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, 45(6), 1321-1324.

Congdon, W. J., Kling, J. R., and Mullainathan, S. (2009). Behavioral Economics and Tax Policy. *National Tax Journal*, 62, 375-386.

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.

Cooter, R. (1998). Expressive Law And Economics. *The Journal of Legal Studies*, 27, (S2), 585-607.

Copeland, P. V., and Cuccia, A. D. (2002). Multiple Determinants of Framing Referents in Tax Reporting and Compliance. *Organizational Behavior and Human Decision Processes*, 88, 499-526.

Coricelli, G., Joly, M., Montmarquette, C., and Villeval, M.C. (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics*, 13, 226-247.

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, 1-18.

Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.

Cullis, J., Jones, P., and Savoia, A. (2012). Social norms and tax compliance: Framing the decision to pay tax. *The Journal of Socio-Economics*, 41, 159-168.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.

De Bartolome, C. A. M. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics*, 56, 79-96.

Deck, C., and Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, 78, 97-119.

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.

Dickinson, D. L. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, 17(4), 638-670.

Djanali, I., and Sheehan-Connor, D. (2012). Tax affinity hypothesis: Do we really hate paying taxes? *Journal of Economic Psychology*, 33 (4), 758-775.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo Reciprocans: Survey Evidence on Behavioural Outcomes. *The Economic Journal*, 119(536), 592-612.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3), 1238-60.

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

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.

Durham, Y., Manly, T.S., and Ritsema, C. (2014). The effects of income source, context, and income level on tax compliance decisions in a dynamic experiment. *Journal of Economic Psychology*, 220–233.

Dwenger, N., Kleven, H., Rasul, I., and Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203-32.

Elffers, H., and Hessing, D. J. (1997). Influencing the prospects of tax evasion. *Journal of Economic Psychology*, 18, 289-304.

Elliot, W., Hodge, F., Kennedy, J., and Pronk, M. (2007). Are M.B.A. Students a Good Proxy for Nonprofessional Investors? *The Accounting Review*, 82, 139-168.

Engström, P., Nordblom, K., Ohlsson, H., and Persson, A. (2015). Tax compliance and loss aversion. *American Economic Journal: Economic Policy*, 7, 132-164.

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

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 Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535-538.

Falk, A., and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5), 1611-1630.

Falsetta, D., Rupert, T. J., and Wright, A. M. (2013). The effect of the timing and direction of capital gain tax changes on investment in risky assets. *The Accounting Review*, 88(2), 499-520.

FDP (2010). Für faire Finanzbeziehungen zwischen Bürger und Staat. Ob als Steuerzahler oder als Transferempfänger – Leistung muss sich immer lohnen. Beschluss des 61. Ordentlichen Bundesparteitags der FDP.

Fehr, E., and Falk, A. (2002). Psychological foundations of incentives. *European economic review*, 46(4-5), 687-724.

Fehr, E., and Goette, L. (2007). Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *American Economic Review*, 97(1), 298-317.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42(1), 1-34.

Feld, L. P., and Frey, B. S. (2002). Trust breeds trust: How taxpayers are treated. *Economics of governance*, 3(2), 87-99.

Feld, L. P., and Frey, B. S. (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law & Policy*, 29(1), 102-120.

Feldman, N. E. (2010). Mental accounting effects of income tax shifting. *The Review of Economics and Statistics*, 92(1), 70-86.

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

Finkelstein, A. (2009). E-ZTAX: Tax Salience and Tax Rates. *The Quarterly Journal of Economics*, 124, 969-1010.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10 (2), 171-178.

Fisher, R. C., Goddeeris, J. H., and Young, J. C. (1989). Participation in tax amnesties: The individual income tax. *National Tax Journal*, 15-27.

Fochmann, M., Hechtner, F., Kirchler, E., and Mohr, P. (2019). When happy people make society unhappy: Incidental emotions affect compliance behavior. *Working Paper*, Available at SSRN 3259071.

Fochmann, M., and Hemmerich, K. (2014). Real Tax Effects and Tax Perception Effects in Decisions on Asset Allocation. *Arqus-Discussion Paper Nr. 156*, April 2014.

Fochmann, M., and Hemmerich, K. (2018). Income Taxes and Risky Investment Decisions: An Experiment on Behavioral Tax Biases. *Journal of Institutional and Theoretical Economics (JITE)*, 174(4), 651-688.

Fochmann, M., and Weimann, J. (2013). The effects of tax salience and tax experience on individual work efforts in a framed field experiment. *FinanzArchiv: Public Finance Analysis*, 69(4), 511-542.

Fochmann, M., Weimann, J., Blaufus, K., Hundsdoerfer, J., and Kiesewetter, D. (2013). Net Wage Illusion in a Real-Effort Experiment. *The Scandinavian Journal of Economics*, 115(2), 476-484.

Fochmann, M. and Wolf, N. (2019). Framing and Salience Effects in Tax Evasion Decisions – An Experiment on Underreporting and Overdeducting. *Journal of Economics Psychology*, 72, 260-277.

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, 136-159.

Fujii, E. T., and Hawley, C. B. (1988). On the Accuracy of Tax Perceptions. *Review of Economics & Statistics*, 70, 344-347.

Gemünden, H. G. (1985). Perceived risk and information search. *A systematic meta-analysis of the empirical evidence. International Journal of Research in Marketing*, 2, 79-100.

Gensemer, B. L., Lean, J. A., and Neenan, W. B. (1965). Awareness of Marginal Income Tax Rates Among High-Income Taxpayers. *National Tax Journal*, 18, 258-267.

Gerhardt, H., Biele, G. P., Heekeren, H. R., and Uhlig, H. (2016). Cognitive Load Increases Risk Aversion. *SFB 649 Discussion Paper*, 2016-011.

Gilovich, T., Griffin, D., and Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *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.

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

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93, 293-300.

Gneezy, U., and Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, 29(1), 1-17.

Graetz, M. J., Reinganum, J. F., and Wilde, L. L. (1986). The tax compliance game: Toward an interactive theory of law enforcement. *Journal of Law, Economics, & Organization*, 2, 1-32.

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

Grolleau, G., Kocher, M. G., and Sutan, A. (2016). Cheating and Loss Aversion: Do People Cheat More to Avoid a Loss? *Management Science*, 62(12), 3428-3438.

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

Harrison, G. W., and List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, 42 (4), 1009-1055.

Hartley, R., and Farrell, L. (2002). Can Expected Utility Theory Explain Gambling? *American Economic Review*, 92(3), 613-24.

Hasegawa, M., Hoopes, J.L., Ishida, R., and Slemrod, J. (2013). The effect of public disclosure on reported taxable income: Evidence from individuals and corporations in Japan. *National Tax Journal*, 66, 571-608.

Hasseldine, J. (1998a). Prospect Theory and Tax Reporting Decisions: Implications for Tax Administrators. *Bulletin for International Fiscal Documentation* 52, 501-505.

Hasseldine, J. (1998b). Tax amnesties: an international review. *Bulletin for International Fiscal Documentation*, 52, 303-310.

Hatfield, R. C., Jackson, S. B., and Schafer, J. K. (2008). An investigation of the relation between tax professionals, tax refunds, and fees. *Behavioral Research in Accounting*, 20, 19-35.

Hausman, J. A. (1981). Labor Supply. Aaron, H., J. and Pechmman, J. A. (1981). How Taxes Affect Economic Behavior. Washington DC: The Brookings Institution, 27-72.

Hayashi, A. T., Nakamura, B. K., and Gamage, D. (2012). Experimental Evidence of Tax Salience and the Labor-Leisure Decision: Anchoring, Tax Aversion, or Complexity? *Public Finance Review*, 41(2), 203-226.

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

Heath, C., and Soll, J. B. (1996). Mental Budgeting and Consumer Decisions. *Journal of Consumer Research*, 23(1), 40-52.

Hechtner, F. (2010). Eine theoretische und empirische Studie über Einkommensteuertarife aus Sicht der Wirtschaftswissenschaft – Progressionswirkungen der synthetischen Einkommensteuer, Schedulenbesteuerung und Vermeidung von Doppelbesteuerungen. Berlin: Doctoral dissertation.

Hedesstrom, T. M., Svedsater, H., and Garling, T. (2004). Identifying heuristic choice rules in the Swedish premium pension scheme. *Journal of Behavioral Finance*, 5(1), 32-42.

Hicks, J. R. (1935). The Theory of Wages. London: MacMillan.

Hitz, J. M., Ernstberger, J., and Stich, M. (2012). Enforcement of accounting standards in Europe: capital-market-based evidence for the two-tier mechanism in Germany. *European Accounting Review*, 21, 253-281.

HM Treasury (2015). Strengthening the incentive to save: a consultation on pensions tax relief (July 2015), URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/442160/Strengthening_the_incentive_to_save_consultation__web_.pdf. accessed 31 January 2020.

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(4), 209-217.

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

Huang, J. (2008). Taxable and tax-deferred investing: a tax-arbitrage approach. *Review of Financial Studies*, 21(5), 2173-2207.

Huck, S., Kübler, D., and Weibull, J. (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization*, 83, 173-185.

Hundsdoerfer, J., and Sichtmann, C. (2007). Zur Gewichtung von Steuern in unternehmerischen Investitionskalkülen: Eine Conjoint-Analyse des Entscheidungsverhaltens selbständiger Ärzte. *Betriebswirtschaftliche Forschung und Praxis*, 604-621.

Hundsdoerfer, J., Sielaff, C., Blaufus, K., Kiesewetter, D., and Weimann, J. (2013). The Influence of Tax Labeling and Earmarking on the Willingness to Contribute - A Conjoint Analysis. *Schmalenbach Business Review*, 359-377.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 5-51.

Internal Revenue Service (2015). Internal Revenue Service Data Book, 2014. Publication 55B, Washington DC.

IRS (2016). Tax Gap Estimates for Tax Years 2008–2010. URL: https:// www.irs.gov/PUP/newsroom /tax%20gap%20estimates%20for%202008%20through%202010.pdf. accessed 31 January 2020.

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

Jackson, S. B., and Hatfield, R. C. (2005). A Note on the Relation between Frames, Perceptions, and Taxpayer Behavior. *Contemporary Accounting Research*, 22, 145-164.

Jackson, S. B., Shoemaker, P. A., Barrick, J. A., and Burton, F. (2005). Taxpayers' prepayment positions and tax return preparation fees. *Contemporary Accounting Research*, 22, 409-447.

Jackson, S. B., and White, R. A. (2008). The effect of tax refunds on taxpayers' willingness to pay higher tax return preparation fees. *Research in Accounting Regulation*, 20, 63-88.

Jamison, J., Karlan, D., and Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization*, 68(3-4), 477-488.

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

John, O. P., and Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of personality and social psychology*, 66(1), 206-219.

Jones, S. R. (1992). Was there a Hawthorne effect? *American Journal of sociology*, 98(3), 451-468.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

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

Kastlunger, B., Kirchler, E., Mittone, L., and Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology*, 30(3), 405-418.

Keane, M. P. (2011). Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, 49 (4), 961-1075.

Kern, M., and Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science*, 20, 378-384.

Killingsworth, M. R. (1983). Labor Supply. Cambridge: Cambridge University Press.

Kimmel, J., and Kniesner, T. J. (1998). New Evidence on Labor Supply: Employment vs. Hours Elasticities by Sex and Marital Status. *Journal of Monetary Economics*, 42, 289-301.

Kirchler, E. (2007). The economic psychology of tax behaviour. Cambridge: 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.

Kirchler, E., and Maciejovsky, B. (2001). Tax compliance within the context of gain and loss situations, expected and current asset position, and profession. *Journal of Economic Psychology*, 22, 173-194.

Kirchler, E., Maciejovsky, B., and Schneider, F. (2003). Everyday representations of tax avoidance, tax evasion, and tax flight: Do legal differences matter? *Journal of Economic Psychology*, 24(4), 535-553.

Kirchler, E., Muehlbacher, S., Hoelzl, E., and Webley, P. (2009). Effort and Aspirations in Tax Evasion: Experimental Evidence. *Applied Psychology*, 58(3), 488-507.

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

Kocher, M. G., Schudy, S., and Spantig, L. (2018). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science*, 64, 3995-4008.

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

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, Hungary, Romania and Russia. *Journal of Economic Psychology*, 34, 169-180.

Korbel, V. (2017). Do we lie in groups? An experimental evidence. *Applied Economics Letters*, 24(15), 1107-1111.

Kugler, T., Kausel, E., and Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 471-482

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.

Langenmayr, D. (2017). Voluntary disclosure of evaded taxes—Increasing revenue, or increasing incentives to evade? *Journal of Public Economics*, 151, 110-125.

Laury, S., and Wallace, S. (2005). Confidentiality and taxpayer compliance. *National Tax Journal*, 58, 427-438.

Lefebvre, M., Pestieau, P., Riedl, A., and Villeval, M. C. (2015). Tax Evasion and Social Information: An Experiment in Belgium, France and the Netherlands. *International Tax and Public Finance*, 22, 401-425.

Levitt, S. D., and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2), 153-174.

Levy, H. (1994). Absolute and Relative Risk Aversion: An Experimental Study. *Journal of Risk and Uncertainty*, 8(3), 289-307.

Liyanarachchi, G. A. (2007). Feasibility of Using Student Subjects in Accounting Experiments: A Review. *Pacific Accounting Review*, 19, 47-67.

López-Laborda, J., and Rodrigo, F. (2003). Tax amnesties and income tax compliance: The case of Spain. *Fiscal Studies*, 24(1), 73-96.

Luitel, H. S., and Sobel, R. S. (2007). The revenue impact of repeated tax amnesties. *Public budgeting & finance*, 27(3), 19-38.

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(6), 678-691.

MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, 58, 593-614.

Malik, A. S., and Schwab, R. M. (1991). The economics of tax amnesties. *Journal of Public Economics*, 46(1), 29-49.

Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.

Markel, D. (2001). Are Shaming Punishments Beautifully Retributive: Retributivism and the Implications for the Alternative Sanctions Debate. *Vanderbilt Law Review*, 54, 2157-2242.

Martinez-Vazquez, J., Harwood, G. B., and Larkins, E. R. (1992). Withholding position and income tax compliance: Some experimental evidence. *Public Finance Quarterly*, 20, 152-174.

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.

Mehrabian, A., and Epstein, N. (1972). A measure of emotional empathy. *Journal of personality*, 40, 525-543.

Mills, L. F., and Sansing, R. C. (2000). Strategic tax and financial reporting decisions: Theory and evidence. *Contemporary Accounting Research*, 17, 85-106.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics*, 35(5), 813-835.

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

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

Pahlke, J., Strasser, S., and Vieider, F. M. (2015). Responsibility effects in decision making under risk. *Journal of Risk and Uncertainty*, 51, 125-146.

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

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

OECD (2019). International Compliance Assurance Programme Pilot Handbook 2.0. Paris: OECD Publishing.

Parle, W. M., and Hirlinger, M. W. (1986). Evaluating the use of tax amnesty by state governments. *Public Administration Review*, 246-255.

Perez-Truglia, R., and Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, 167, 120-137.

Pickhardt, M., and Prinz, A. (2014). Behavioral dynamics of tax evasion–A survey. *Journal of Economic Psychology*, 40, 1-19.

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.

Rabin, M. (1998). Psychology and economics. *Journal of economic literature*, *36*(1), 11-46.

Ramsey, Frank P. (1927). A Contribution to the Theory of Taxation. *Economic Journal*, 37 (145), 47-61.

Randall, D. M., and Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of business ethics*, 10(11), 805-817.

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

Read, D., and Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1(1), 34-49.

Rechberger, S., Hartner, M., Kirchler, E., and Hämmerle, F. K. (2010). Tax amnesties, justice perceptions, and filing behavior: A simulation study. *Law & policy*, 32(2), 214-225.

Rees-Jones, A. (2018). Quantifying Loss-Averse Tax Manipulation. *Review of Economic Studies*, 85, 1251-1278.

Reinganum, J. F., and Wilde, L. L. (1986). Equilibrium verification and reporting policies in a model of tax compliance. *International Economic Review*, 739-760.

Remus, W. (1996). Will Behavioral Research on Managerial Decision Making Generalize to Managers? *Managerial and Decision Economics*, 17, 93-101.

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.

Riedl, A., and Tyran, J. R. (2005). Tax liability side equivalence in gift-exchange labor markets. *Journal of public Economics*, 89(11/12), 2369-2382.

Ritov, I., and Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and human decision processes*, 64, 119-127.

Robben, H., Webley, P., Elffers, H., and Hessing, D. (1990a). Decision Frames, Opportunity, and Tax Evasion: An Experimental Approach. *Journal of Economic Behavior and Organization*, 14, 353-361.

Robben, H., Webley, P., Weigel, R., Wärneryd, K., Kinsey, K., Hessing, D., Martin, F., Elffers, H., Wahlund, R., Van Langenhove, L., Long, S., and Scholz, J. (1990b). Decision Frames and Opportunity as Determinants of Tax Cheating. *Journal of Economic Psychology*, 11, 341-364.

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

Roethlisberger, F. J., and Dickson, W. J. (1939). Management and the Worker. Cambridge: Harvard University Press.

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.

Rupert, T. J., and Fischer, C. M. (1995). An Empirical Investigation of Taxpayer Awareness of Marginal Tax Rates. *Journal of the American Taxation Association*, 17, 36-59.

Rupert, T. J., Single, L. E., and Wright, A. M. (2003). The Impact of Floors and Phase-Outs on Taxpayers Decisions and Understanding of Marginal Tax Rates. *Journal of the American Taxation Association*, 25, 72-86.

Rupert, T. J., and Wright, A. M. (1998). The Use of Marginal Tax Rates in Decision Making: The Impact of Tax Rate Visibility. *The Journal of the American Taxation Association*, 20, 83-99.

Sansing, R. C. (1993). Information acquisition in a tax compliance game. *Accounting Review*, 874-884.

Sausgruber, R., and Tyran, J. R. (2005). Testing the Mill hypothesis of fiscal illusion. *Public Choice*, 122(1-2), 39-68.

Sausgruber, R., and Tyran, J. R. (2011). Are we taxing ourselves? How deliberation and experience shape voting on taxes. *Journal of Public Economics*, 95, 164-176.

Schepanski, A., and Shearer, T. (1995). A prospect theory account of the income tax withholding phenomenon. *Organizational Behavior and Human Decision Processes*, 63, 174-186.

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.

Sillamaa, M.-A. (1999). Taxpayer behavior in response to taxation: comment and new experimental evidence. *Journal of Accounting and Public Policy*, 18, 165-177.

Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, 59 (4), S209-S224.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *The Journal of Economic Perspectives*, 21, 25-48.

Slemrod, J. (2019). Tax Compliance and Enforcement. *Journal of Economic Literature*, 57 (4), 904-954.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2), 274-279.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.

Sobel, M. E. (1986). Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models. *Sociological Methodology*, 16, 159-186.

Sobel, M. E. (1987). Direct and Indirect Effects in Linear Structural Equation Models. *Sociological Methods & Research*, 16(1), 155-176.

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

Stata Corp. (2015). Stata Structural Equation Modeling Reference Manual Release 14. Texas: Stata Press, Stata Corp. LP.

Stella, P. (1991). An economic analysis of tax amnesties. *Journal of Public Economics*, 46(3), 383-400.

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.

Sutter, M., and Weck-Hannemann, H. (2003). Taxation and the Veil of Ignorance–A real effort experiment on the Laffer curve. *Public Choice*, 115(1-2), 217-240.

Swenson, C. W. (1988). Taxpayers behavior in response to taxation: An experimental analysis. *Journal of Accounting and Public Policy*, 7, 1-28.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1), 39-60.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214.

Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *Journal of Economic Perspectives*, 4(1), 193-205.

Thaler, R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12(3), 183-206.

Thaler, R. H., and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of political Economy*, 89(2), 392-406.

The Nobel Foundation (2002). The Prize in Economic Sciences 2002. URL: https://www.nobelprize.org/prizes/economic-sciences/2002/summary/. accessed 31 January 2020.

The Nobel Foundation (2017). The Prize in Economic Sciences 2017. URL: https://www.nobelprize.org/prizes/economic-sciences/2017/summary/. accessed 31 January 2020.

The Nobel Foundation (2019). The Prize in Economic Sciences 2019. URL: https://www.nobelprize.org/prizes/economic-sciences/2019/summary/. accessed 31 January 2020.

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

Torgler, B. (2013). A field experiment in moral suasion and tax compliance focusing on underdeclaration and overdeduction. *FinanzArchiv: Public Finance Analysis*, 69, 393-411.

Torgler, B., Demir, I. C., Macintyre, A., and Schaffner, M. (2008). Causes and Consequences of Tax Morale: An Empirical Investigation. *Economic Analysis & Policy*, 38, 313-339.

Torgler, B., and Schaltegger, C. A. (2005). Tax amnesties and political participation. *Public Finance Review*, 33(3), 403-431.

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

Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185 (4157), 1124-1131.

Vihanto, M. (2003). Tax evasion and the psychology of the social contract. *The Journal of Socio-Economics*, 32(2), 111-125.

Warren Jr, A. C. (1986). The Timing of Taxes. *National Tax Journal*, 39(4), 499-505.

Wartick, M.L., Madeo, S.A., and Vines, C.C. (1999). Reward dominance in tax-reporting experiments: The role of context. *Journal of the American Taxation Association*, 21, 20-31.

Weber, M., and Schram, A. (2017). The Non-equivalence of Labour Market Taxes: A Real-effort Experiment. *The Economic Journal*, 127(604), 2187-2215.

Weber, M., and Zuchel, H. (2005). How do prior outcomes affect risk attitude? Comparing escalation of commitment and the house-money effect. *Decision Analysis*, 2(1), 30-43.

Weimann, J. and Brosig-Koch, J. (2019). Methods in Experimental Economics: An Introduction. Wiesbaden: Springer.

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.

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, 26(6), 862-883.

Whitney, P., Rinehart, C. A., and Hinson, J. M. (2008). Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic Bulletin & Review*, 15, 1179-1184.

Whyte, G. (1993). Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes*, 54, 430-455.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1, 80-83.

Wilson, J.Q., and Kelling, G.L. (1982). Broken windows. *Atlantic Monthly*, 249, 29-38.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115, 157-168.

Yaniv, G. (1999). Tax compliance and advance tax payments: A prospect theory analysis. *National Tax Journal*, 52, 753-764.

Yoo, K. Y., and De Serres, A. (2005). Tax treatment of private pension savings in OECD countries. *OECD Economic Studies*, 39, 73-110.

Ziliak, J. P., and Kniesner, T. J. (1999). Estimating Life-Cycle Labor Supply Tax Effects. *Journal of Political Economy*, 107, 326-359.