



universität  
wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Wahrgenommene Servicequalität und Steuerehrlichkeit -  
Eine Prozessanalyse mit MouselabWEB“

verfasst von / submitted by

Amanda Wöhler, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of  
Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Masterstudium Psychologie  
(Schwerpunkt: Arbeit, Wirtschaft und  
Gesellschaft)

Betreut von / Supervisor:

Univ.-Prof. Dr. Erich Kirchler

## Zusammenfassung

Eine höhere Servicequalität (SQ) wird assoziiert mit einem höheren Vertrauen in die Steuerbehörde und einer höheren Steuerehrlichkeit. Falls Steuerehrlichkeit durch dynamische Interaktionen zwischen wahrgenommener Macht der Steuerbehörde und Vertrauen in diese, entstehe, sollte bei niedrigerer SQ mehr Aufmerksamkeit auf den Strafparametern liegen. In der vorliegenden Studie wurde die SQ einer fiktiven Steuerbehörde manipuliert und die zugrunde liegenden Informationsverarbeitungsprozesse mittels MouselabWEB aufgezeichnet. In 18 Entscheidungsrunden wurden von 485 Teilnehmer\*innen, unter Variation von vier Parametern (Einkommen, Steuerrate, Prüfwahrscheinlichkeit und Höhe der Strafe), inzentivierte Steuerentscheidungen getroffen. Es zeigte sich bei niedriger SQ weder eine signifikant höhere Steuerehrlichkeit, noch eine längere Betrachtungsdauer oder -häufigkeit der Strafparameter. Die wahrgenommene Service Orientierung ist bei hoher SQ zwar signifikant höher, liegt jedoch auch bei niedriger SQ noch im mittleren Bereich. Die Ergebnisse legen nahe, dass die Service Qualität in mittlerer Ausprägung nicht ausreicht, die hypothetisierten Effekte zu zeigen. Dies könnte in zukünftigen Studien mit einer stärker variierten SQ weiter untersucht werden.

## Abstract

Higher service quality (SQ) is associated with higher tax honesty and trust in the tax authority. If tax honesty arises from dynamic interactions between perceived power of the tax authority and trust in it, more attention should be paid to penalty parameters when SQ is lower. In the present study, the SQ of a fictitious tax authority was manipulated and the underlying information processing procedures were recorded using MouselabWEB. In 18 decision rounds, 485 participants made incentivized tax decisions while varying four parameters (income, tax rate, audit probability, and fine rate). Neither significantly higher tax honesty nor longer viewing duration or frequency of penalty parameters was found at low SQ. Perceived service orientation is significantly higher at high SQ, but is still in the middle range even at low SQ. The results suggest that service quality at medium levels is not sufficient to show the hypothesized effects. This could be further investigated in future studies with a more varied SQ.

## Einleitung

Ein Staat ist, neben erhobenen Gebühren und Abgaben, auf die Zahlung von Steuern angewiesen, um seine Aufgaben zu erfüllen und die Bürger\*innen sind gesetzlich verpflichtet diese regelkonform zu bezahlen. Österreich hat im Jahr 2020 alleine aus der Einkommens- und Vermögenssteuer 48.341.000.000€ eingenommen (Statistik Austria, 2022). Aber wie kann der Staat die Einhaltung der Steuergesetze möglichst effizient gewährleisten?

Das *klassische ökonomische Standardmodell* von Allingham und Sandmo aus dem Jahr 1972 umfasst vier Parameter, die der Entscheidung Steuern zu bezahlen oder sie nicht zu bezahlen zu Grunde liegen. Die Höhe des Einkommens, die Höhe der Steuerrate, die Wahrscheinlichkeit geprüft zu werden und die Höhe der Strafe. Aufgrund dieser Parameter wird angenommen, dass Steuerzahler\*innen den erwarteten Nutzen bei Steuerhinterziehung berechnen und diesen mit dem Nettobetrag bei steuerkonformem Verhalten vergleichen. Personen zahlen nach diesem Modell lediglich aufgrund der ökonomischen Konsequenzen einer möglichen Prüfung und Bestrafung ihre Steuern (Alm, 2012). Allerdings zeigen sich in Studien diesbezüglich niedrigere und teilweise widersprüchlichere Effekte (Fischer et al., 1992; Kirchler et al., 2008; Alm et al., 2012). Die Steuerehrlichkeit ist in den meisten Ländern trotz niedriger Prüfwahrscheinlichkeiten und Strafen hoch (Alm et al., 1992). Beispielsweise berichtete das IRS (Internal Revenue Service) für das Jahr 2015 lediglich eine Gesamtprüfungsquote von 0.8 Prozent (Internal Revenue Service, 2016, zitiert nach Alm, 2019). Würden Steuerzahler\*innen alleine aus ökonomischen Gesichtspunkten entscheiden, sollte die Quote der Steuerehrlichkeit also deutlich niedriger ausfallen. Somit rücken andere Faktoren in den Fokus der wissenschaftlichen Auseinandersetzung mit Steuerehrlichkeit, wie zum Beispiel Vertrauen in die Steuerbehörde, Prozedurale Gerechtigkeit und Servicequalität (Gangl et al., 2013). Auch diese zeigen jedoch widersprüchliche Ergebnisse und lassen somit ein Zusammenspiel verschiedener Faktoren vermuten.

Das *Slippery Slope Modell* (Kirchler et al., 2008) postuliert zwei relevante Dimensionen, die die Einstellungen zum Steuerzahlen ausdrücken. Zum einen die Macht der Steuerbehörden und zum anderen das Vertrauen in die Steuerbehörden. Es vereint somit die verschiedenen Ansätze in einem gemeinsamen Modell, nach dem verschiedene

Motive zu Steuerehrlichkeit führen können. Entscheidungen sind somit als multifaktorielles Produkt von ökonomischen und sozialpsychologischen Einflüssen zu sehen (Mittone, 2005). Zwischen Macht und Vertrauen werden dynamische Interaktionen angenommen und das Steuerklima variiert auf einem Kontinuum zwischen antagonistischem und synergetischem Klima. Ein antagonistisches Klima beschreibt hierbei eine Situation, in der Steuerbehörden und Steuerzahler\*innen gegeneinander arbeiten und die Steuerzahler\*innen sich von den Behörden verfolgt fühlen (Kirchler et al., 2008). Dies kann dazu führen, dass diese kein schlechtes Gewissen bei dem Versuch haben, ihr Einkommen vor den Behörden zu verstecken (Braithwaite, 2003a). Die Distanz zu den Behörden ist häufig groß, wenig respektvoll und es gibt wenig positive Gefühle der Steuerbehörde gegenüber. Die freiwillige Steuerehrlichkeit sinkt und es wird eher eine rationale ökonomische Abwägung getroffen, ob es sich lohnt Steuern zu zahlen (Kirchler et al., 2008). In einem synergetischen Klima arbeiten die Steuerzahler\*innen und die Steuerbehörden, welche als serviceorientierte Teilnehmer der Gemeinschaft wahrgenommen werden, zusammen. Die Behörden bemühen sich um transparente Prozesse und ein respektvolles und unterstützendes Verfahren der Steuererhebung, in dem die Steuerzahler\*innen sich als Kund\*innen fühlen können. Die soziale Distanz ist eher gering und freiwillige Einhaltung der Steuerregeln ist wahrscheinlicher, Personen fühlen sich verpflichtet ihren Teil beizutragen (Kirchler et al., 2008). Eine Fallstudie über die Gestaltung des Steuersystems in der Schweiz zeigt die Bedeutung einer freundlichen und respektvollen Behandlung, neben der Wertschätzung der Infrastruktur und politischen Faktoren, für die Erhöhung der freiwilligen Einhaltung der Steuerregeln (Feld & Frey, 2005). Je mehr Fälle eine Steuerbehörde überprüft, desto mehr Ressourcen muss sie dafür aufwenden. Zum einen müssen Steuerbehörden ihre Ressourcen erhöhen, je mehr Fälle sie überprüfen und je mehr Fälle der Steuerhinterziehung sie aufdecken. Die Fälle müssen von Steuerbeamt\*innen untersucht werden um die Schwere der Vergehen zu ermitteln und Strafen festzusetzen. Eine erhöhte freiwillige Steuerehrlichkeit trägt also dazu bei, diesen Aufwand zu reduzieren.

## Servicequalität

Prozedurale Gerechtigkeit beschreibt die wahrgenommene Gerechtigkeit von Verfahren, welche zur Entscheidungsfindung beitragen und die wahrgenommene Behandlung, welche man von Entscheidungsträger\*innen, hier der Steuerbehörde, erhält (Murphy, 2003; Hartner et al., 2008; Farrar, 2015; van Dijke, 2019). Sie umfasst auch die Servicequalität (SQ), welche beschreibt, wie die Steuerbehörde mit den Steuerzahler\*innen umgeht. Um eine höhere Steuerehrlichkeit zu erzielen, sollte die Behörde die Steuerzahler\*innen mit Respekt behandeln, unterstützenden Service bereitstellen und Teilhabe am Prozess ermöglichen (Alm & Torgler, 2011; Braithwaite, 2003b; Kirchler et al., 2008). Gangl et al. (S.503) konnten 2013 außerdem einen hohen Zusammenhang zwischen der wahrgenommenen Serviceorientierung und der wahrgenommenen Vertrauenswürdigkeit des Steuerbüros belegen ( $r = 0.74, p < .001$ ). Sowohl das wahrgenommenen Vertrauen als auch die wahrgenommene Serviceorientierung zeigten allerdings kleine Effekte ( $r = 0.13, p < .05$  und  $r = 0.12, p < .01$ ) auf die Intention, die Steuern ordnungsgemäß zu zahlen.

Auch eine Fallstudie aus Singapur belegt den positiven Effekt von Servicequalität auf die Steuerehrlichkeit (Alm & Torgler, 2011). Die Steuererhebungsprozesse wurden reformiert, im Wesentlichen durch Verbesserungen der Servicequalität, indem die Umstellung von einem papiergebundenen Ablagesystem auf ein elektronisches System durchgeführt wurde und darüber hinaus ein online-Service für Fragen jeglicher Art eingeführt wurde. Darüber hinaus konnten die Steuerzahler das gesamte Steuerformular jederzeit einsehen und auch die Einstellung der Beamten gegenüber den Steuerzahldern hat sich verbessert. „Die Steuerverwaltung Singapurs hat sich von der am niedrigsten bewerteten Regierungsbehörde in Bezug auf die öffentliche Zufriedenheit zu einer Behörde entwickelt, die von 90 % der Steuerzahler als zuvorkommend, kompetent und bequem empfunden wird.“ (S. 649, Alm & Torgler, 2011).

## Prozessanalysen mit MouselabWEB

In der vorliegenden Arbeit werden neben der Analyse des Outcomes, also der Steuerehrlichkeit, auch Prozessdaten analysiert, welche eine spezifischere Analyse der Prozesse ermöglichen. Prozessdaten liefern eine Grundlage für eine tiefer gehende

Analyse der Verarbeitungsstrategien der Teilnehmer\*innen bei den Steuerentscheidungen und geben darüber hinaus auch die Möglichkeit, zugrundeliegende Annahmen von Modellen zu überprüfen (Willemse & Johnson, 2019). Steuerexperimente beschränken sich zumeist auf die Erhebung der tatsächlichen Steuerentscheidung und vernachlässigen die zugrundeliegenden kognitiven Prozesse (Kogler et al., 2022). Die Wahrnehmung der dargestellten vier Parameter, nach denen die Steuerentscheidung getroffen wird, lässt sich mit MouselabWEB (Willemse & Johnson, 2019). MouselabWEB ist ein Programm zur Prozessverfolgung, welches einzelne Informationen hinter Boxen versteckt und nur während der Berührung mit der Maus anzeigt. Die Mausbewegungen der Teilnehmer\*innen werden aufgezeichnet und somit auch die Dauer und Häufigkeit der Boxöffnungen und die Übergänge zwischen den Boxen. Da ein Entscheidungsträger eine Information erworben haben muss, um sie zu verwenden und dieser Erwerb zeitlich nah an der Verwendung liegt (Costa-Gomes, Crawford & Broseta, 2001), sollte eine höhere Aufmerksamkeit auf den Parametern liegen, die stärker in die Steuerentscheidung mit einbezogen werden. Da die Teilnehmer\*innen die Boxen öffnen müssen, wenn sie den jeweiligen Wert des darunter befindlichen Parameters sehen möchten, kann somit die Aufmerksamkeit auf der jeweiligen Information messbar gemacht werden (Willemse & Johnson, 2019). Durch eine Prozessanalyse der mit MouselabWEB gesammelten Daten kann die Aufmerksamkeit auf den Parametern untersucht werden und zu einem besseren Verständnis der Heterogenität der Steuerzahler\*innen, beitragen (Willemse & Johnson, 2019).

## **Fragestellung**

Die vorliegende Arbeit untersucht, ob der Einfluss von Prüfwahrscheinlichkeit und Höhe der Strafe, also den mit Macht assoziierten Steuerparametern, auf die Steuerehrlichkeit, von der Verfügbarkeit von soziopsychologischen Informationen, hier der Servicequalität (SQ), abhängt. Es wurde eine Präregistrierung der Hypothesen durchgeführt (abrufbar unter <https://osf.io/ugdt6>). Hypothese eins und zwei beziehen sich dabei auf die Ebene des Outcome, also die Steuerehrlichkeit, Hypothese drei, vier und fünf beziehen sich auf der Prozessebene auf die Betrachtung der mit Macht assoziierten Parameter.

## ***Outcome-Hypothesen***

Eine hohe Servicequalität (SQ) zeigt positive Zusammenhänge mit dem Vertrauen in Autoritäten (Gangl et al., 2013) und mit Steuerehrlichkeit (Alm & Torgler, 2011; Gangl et al., 2013). Dies führt zur folgenden Hypothese:

H1: Die relative Steuerehrlichkeit ist bei hoher SQ höher als bei niedriger SQ.

Da laut Slippery Slope Modell (Kirchler et al., 2008) bei niedrigem Vertrauen in die Steuerbehörde, der Einfluss der wahrgenommenen Macht auf die Steuerehrlichkeit höher sein sollte, ergibt sich hieraus die folgende Hypothese:

H2: Veränderungen der Strafparameter führen bei niedriger SQ zu einer stärkeren behavioralen Reaktion als bei hoher SQ.

## ***Prozess-Hypothesen***

Die drei Prozesshypothesen liegen dem Prozess hinter Hypothese 2 zugrunde. Das Slippery Slope Modell (Kirchler, Hoelzl, & Wahl, 2008) postuliert eine dynamische Interaktion zwischen Vertrauen in die Steuerbehörde und wahrgenommener Macht der Steuerbehörde. Nach dieser These sollte bei einem geringeren Vertrauen in die Steuerbehörde, die wahrgenommene Macht einen höheren Einfluss auf die Steuerentscheidung haben bzw. bei der Steuererhebung stärker in die Entscheidung mit einbezogen werden, als bei höherem Vertrauen. Die beiden Parameter die bei der experimentellen Erhebung die wahrgenommene Macht der Steuerbehörde beeinflussen, sind die Strafhöhe und die Prüfwahrscheinlichkeit. Ist die Wahrscheinlichkeit, dass die abgegebene Steuererklärung geprüft wird, höher, bzw. fällt bei einer entdeckten Steuerhinterziehung die Strafe höher aus, wird der Staat mächtiger wahrgenommen, da sein Potential Steuerhinterziehungen aufzudecken, ebenfalls höher ist (Kirchler et al., 2008). Hat die wahrgenommene Macht bei niedriger SQ einen höheren Einfluss auf die Steuerehrlichkeit, sollte auch mehr Aufmerksamkeit auf den Strafparametern liegen. Aufmerksamkeit wird in den drei Hypothesen operationalisiert durch Betrachtungsdauer, Betrachtungshäufigkeit, Übergänge zwischen den Strafparametern und längere Entscheidungszeiten.

H3: Teilnehmer\*innen der hohen SQ-Bedingung schenken den Strafparametern weniger Aufmerksamkeit als Teilnehmer\*innen der niedrigen SQ-Bedingung. Dies zeigt sich in weniger und kürzeren Öffnungen der Boxen der Prüfwahrscheinlichkeit und der Strafhöhe bei hoher SQ im Vergleich zu niedriger SQ.

H4: Teilnehmer\*innen der niedrigen SQ-Bedingung haben mehr Übergänge zwischen Prüfwahrscheinlichkeit und Strafhöhe im Vergleich zu Teilnehmer\*innen der hohen SQ-Bedingung.

H5: Teilnehmer\*innen der niedrigen SQ-Bedingung haben längere Entscheidungszeiten pro Runde im Vergleich zu Teilnehmer\*innen der hohen SQ-Bedingung.

## **Methode**

### **Stichprobe**

Es wurde eine Stichprobe von  $N = 497$  über prolific erhoben. Voraussetzungen der Teilnahme waren ein Wohnsitz in der UK und eine study approval rate auf prolific von mindestens 96/100. Die Erhebung wurde auf Englisch durchgeführt, die Teilnehmer\*innen gaben ihre Englischkenntnisse durchschnittlich auf einer siebenstufigen Likertskaala mit 6.85 ( $SD = 0.42$ ) an, alle Teilnehmer\*innen liegen zwischen vier und sieben. Insgesamt wurden elf Teilnehmer\*innen ausgeschlossen, zwei aufgrund technischer Schwierigkeiten und neun aufgrund fehlender Steuerehrlichkeitsdaten von mehr als vier Runden. Die verbleibende Stichprobe besteht aus  $N = 485$  Teilnehmer\*innen (274 Frauen, 206 Männer) mit einem Durchschnittsalter von 35.3 Jahren ( $SD = 12.09$ ), das zwischen 18 und 75 Jahren liegt. Im Vorfeld der Studie wurde eine Sensitivitätsanalyse mit G\*Power (Faul et al., 2007) für eine ANOVA (fixed effects, special, main effects, interactions) durchgeführt und diese ergab, dass bei einer Stichprobengröße von  $N = 288$  ein kleiner Effekt  $f = 0.22$  mit einer Power von  $1-\beta = 0.8$  gefunden werden kann ( $\alpha = 0.05$ ).

## **Versuchspran**

Um den Faktor der Servicequalität zu manipulieren, wurden die Teilnehmer\*innen zu Beginn randomisiert einer von zwei Bedingungen zugeteilt, der hohen oder der niedrigen SQ. Um die Manipulation in den beiden Bedingungen zu testen, wurde ein Pretest mit hoher und niedriger SQ durchgeführt.

In der Bedingung der hohen SQ wurde die Steuerbehörde in der Einleitung als Prozess erleichternd und hilfsbereit beschrieben. Außerdem gebe sie die Möglichkeit am Ende ein Feedback über den Steuererhebungsprozess abzugeben. Im Rahmen der Steuererhebungen erläutert die Steuerbehörde im Brief den Vorgang, dass der Identifikationscode automatisiert im darauf folgenden Formular leicht modifiziert eingetragen wird. Die Teilnehmer\*innen mussten diesen lediglich bestätigen.

We already performed this modification for you and prefilled the modified code on the next page of this form. We hope to be able to simplify the filing process for you by providing this service.

In der Bedingung der niedrigen SQ wurde die Steuerbehörde als den Prozess verkomplizierend und nicht hilfsbereit beschrieben und darüber hinaus gebe es keine Möglichkeit für Feedback. Die Teilnehmer\*innen mussten den Identifikationscode selbst modifizieren, indem sie alle Bindestriche entfernen. Darüber hinaus wurden sie im Brief darauf hingewiesen, dass die Steuerbehörde diesen Prozess auch in Zukunft nicht vereinfachen möchte und Feedback diesbezüglich nicht erwünscht sei:

It is mandatory for you to make the demanded changes to your identification number. It is not possible to avoid this process. We do not see it as a priority for the tax administration to simplify this process in the future.

Der Identifikationscode konnte kopiert und eingefügt werden, außerdem konnten die Teilnehmer\*innen das Steuerformular mit dem Identifikationscode erneut aufrufen.

## **Steuerehrlichkeit**

Die Steuerehrlichkeit wurde auf der behavioralen Ebene als relativer Anteil der gezahlten Steuer erhoben. Die Ausprägungen von vier within-Faktoren wurden in 18 Runden variiert: Höhe des Einkommens (1000-2000 Experimental Currency Units),

Steuerrate (20% vs. 30% vs. 40%), Prüfwahrscheinlichkeit (5% vs. 10% vs. 15%) und Strafhöhe (hinterzogener Betrag plus 50% Strafe vs. plus 100% Strafe). Die Reihenfolge der Runden erfolgte für alle Teilnehmer\*innen randomisiert. Die Darstellung der vier Faktoren erfolgte durch MouselabWEB. In diesem Prozessverfolgungsprogramm wird der Wert der Parameter ausgeblendet, solange sich die Computermaus nicht auf der jeweiligen Box befindet (s. Abb.1). Die Teilnehmer\*innen erhielten randomisiert zwei verschiedene Anordnungen der Felder, diese blieb über das Experiment hin konstant.

### Abbildung 1

*Darstellung der Steuerparameter in MouselabWEB. Links sind alle Boxen geschlossen, rechts befindet sich die Maus auf der Box der Steuerhöhe.*



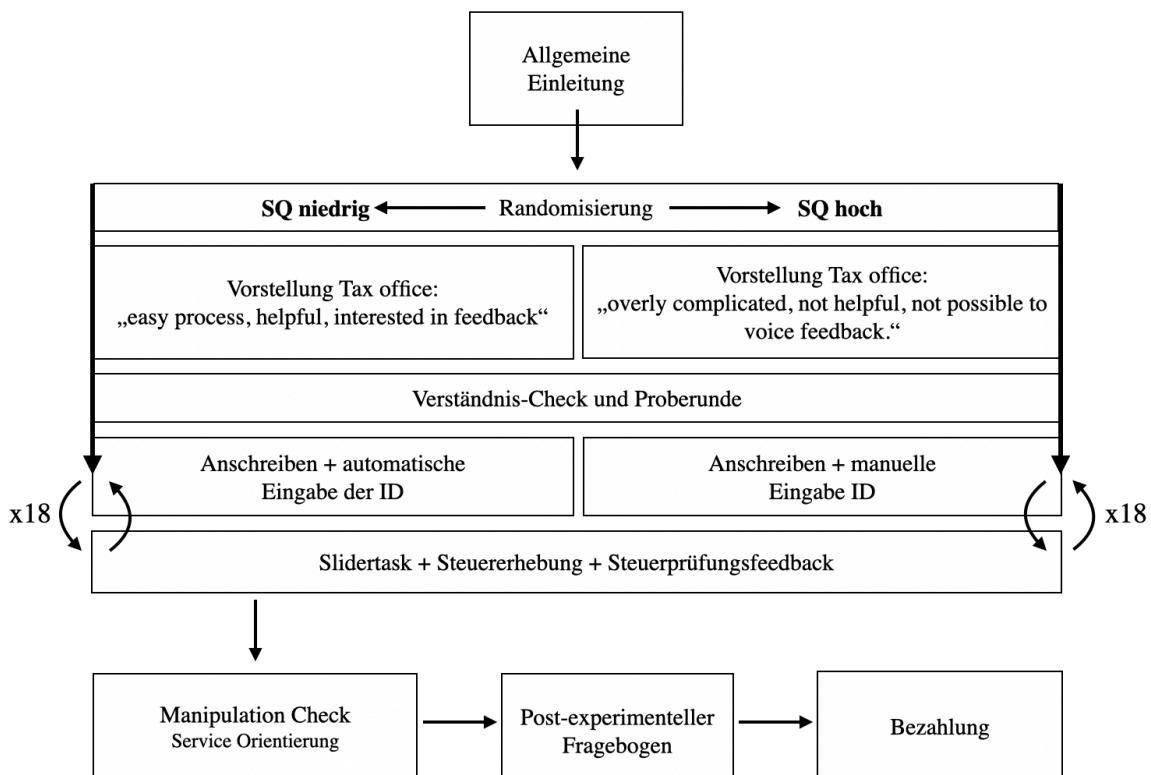
### Erhebung

Die online durchgeführte Erhebung umfasst einen Abschnitt mit wiederholten Messungen der Steuerehrlichkeit, Manipulation-Checks und einen postexperimentellen Fragebogen (s. Abb. 2), ein Link zur vollständigen Darstellung der Erhebung befindet sich in Anhang A. Die Teilnehmer\*innen erhalten eine Teilnahmeentschädigung von mindestens £2.10 und das Nettoeinkommen einer zufälligen Steuerrunde (£0.40 - £4.00).

Vor der ersten echten Steuerrunde wurden die vier Faktoren, die Schieberegleraufgabe und der Ablauf der Steuererhebung erklärt und es fand eine Proberunde statt. Außerdem wurde die Steuerbehörde in den beiden Bedingungen verschieden vorgestellt. Jede Runde begann mit einem Effort Task, bei der zehn Schieberegler innerhalb von 20 Sekunden auf einer Skala von 0-100 genau mittig positioniert werden sollen, es wurde lediglich der genaue Wert von 50 als korrekt gewertet. Hiermit konnten die Teilnehmer\*innen zusätzlich zu ihrem festen Einkommen von 1000 ECU (Experimental Currency Units, 1€ = 80 ECU) je Runde, 100 ECU je Schieberegler verdienen und das Einkommen je Runde betrug somit 1000 bis 2000 ECU. Nach dem Effort Task erhielten die Teilnehmer\*innen einen Brief der Steuerbehörde, in dem sie einen 12-stelligen Identifikationscode zugewiesen bekamen, welcher durch Bindestriche in drei Blöcke eingeteilt war. Die Präsentation der Materialien ähnelt einem echten Papier-Briefdesign um ein möglichst realistisches Erscheinungsbild eines traditionellen Steuerformulars zu erzeugen. Auch der Inhalt des Briefs ist kontextreich formuliert.

## Abbildung 2

Flow Chart des Erhebungsablaufs.



Anschließend folgte in beiden Bedingungen die Steuererhebung über 18 Runden in MouselabWEB, bei der das Einkommen (ECU), die Steuerrate (Prozent) und die Höhe der zu zahlenden Steuer (ECU), die Prüfwahrscheinlichkeit (Prozent) und die Strafhöhe (ECU) eingeblendet wurde. Die Teilnehmer\*innen gaben an, welchen Betrag sie bezahlen und konnten das Formular dann abgeben. Auf der nächsten Seite erhielten sie ein Feedback, ob sie in dieser Runde einer Steuerprüfung unterzogen wurden und welches Einkommen, nach einer möglichen Strafe und abzüglich der bezahlten Steuern, sie in dieser Runde erspielt hatten. Nach den Steuerrunden hatten die Teilnehmer\*innen der hohen SQ die Möglichkeit, Feedback bezüglich ihrer Wahrnehmung des Prozesses zu geben.

### ***Manipulation-Checks***

Im ersten prä-experimentellen Manipulation-Check wurde beispielhaft im MouselabWEB-Design, aber ohne ausgeblendete Werte, eine Steuerrunde dargestellt. Vier Fragen prüften das Verständnis der Instruktionen der vier Steuerparameter und der Berechnung des Nettoeinkommens (z.B. „How much would your remaining income be if you pay 200 ECU and there is no audit?“). Auf der nächsten Seite erhielten die Teilnehmer\*innen die richtigen Lösungen und farbig markiert (rot = falsch; grün = richtig) ihre eigenen Antworten.

Nach den Steuererhebungsrounden fand ein Attention-Check statt, die Teilnehmer\*innen erhielten eine Multiple-Choice-Frage: „Think back of the instructions of the study. How was the Central Tax Office described?“ mit vier Antwortmöglichkeiten. Die Teilnehmer\*innen der beiden Bedingungen hatten jeweils eine mögliche richtige Antwort (niedrige SQ: „...helpful towards the taxpayers in their filing process.“; hohe SQ: „...making the tax filing process overly complicated.“). Darüber hinaus gaben die Teilnehmer\*innen in vier Fragen an, wie aufmerksam und konzentriert sie während der Erhebung waren (z.B. „Did you understand all the information?“; Likert-Skala von 0 = *No, not at all* bis 4 = *Yes, completely*).

Es folgte der Manipulation-Check der wahrgenommenen Serviceorientierung, erhoben nach Gangl et al. (2013). Die Teilnehmer\*innen gaben ihre Zustimmung zu Aussagen bezüglich des Services der Steuerbehörde an (Likert-Skala von 1 = *completely*

*disagree* bis 5 = *completely agree*), z.B.: „With regard to the tax collection in the study, the Central Tax Office does everything possible to serve people“. Neben der wahrgenommenen Service Orientierung wurde auf der nächsten Seite mit erhoben, wie die Teilnehmer\*innen den Prozess rund um den Identifikationscode wahrgenommen haben: „Please think back of the process of receiving and reviewing the tax identification number. Please state your agreement with the following statements.“ Die Teilnehmer\*innen (Likert-Skala von 0 = *Fully disagree* bis 4 = *Fully agree*) gaben ihre Zustimmung zu vier Aussagen diesbezüglich an (z.B. „The process was annoying.“).

### ***Postexperimenteller Fragebogen***

Der postexperimentelle Fragebogen umfasst eine Selbsteinschätzung der Wichtigkeit der vier Steuerparameter für die Teilnehmer\*innen (Schieberegler von *not important at all* bis *very important*): „How important was the following information for your decision?“, des Weiteren sechs Fragen zur Intention der Steuer(un)ehrlichkeit aus dem TAX-I (tax compliance inventory) nach Kirchler & Wahl (2010), z.B.: „When I paid my taxes in the experiment as required by the regulations, I did so because I felt a moral obligation to pay my tax.“ (Likert-Skala von 0 = *Completely disagree* bis 4 = *Completely agree*). Darüber hinaus wurde die Risikoneigung mit einer Frage nach Dohmen et al. (2011) erfasst: „How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?“ (Likert-Skala von 0 = *Definitely avoiding to take risks* bis 9 = *Fully prepared to take risks*). Die Teilnehmer\*innen der hohen SQ wurden nach ihrem Vertrauen in den automatisch übertragenen und umkodierten Identifikationscode gefragt: „During the tax filing you were asked to confirm that the pre-filed code was correct. Did you think the code was always pre-filed correctly?“ (*Yes/No*). Darüber hinaus gaben sie an, ob sie den *go back to information button* benutzt haben (*Yes/No*). Wurde hier *Yes* gewählt, wurden die Teilnehmer\*innen zusätzlich gefragt, wie häufig sie diesen benutzt haben (*1 time only, 2-5 times, 6-12 times, 13-18 times*). Danach wurde das Alter in Jahren, das Geschlecht (*Male, Female, Other*), die Englischkenntnisse (Likert-Skala von 0 = *very low* bis 6 = *very high*) und bisherige Teilnahmen an Steuerehrlichkeitsstudien (*Yes/No*) erhoben. Es wurden zwei Fragen zum technischen Gerät (*Computer, Laptop, Phone, Tablet*) und Eingabegerät

(*Computer mouse, trackpad, Touchscreen, Stylus pen*), an dem teilgenommen wurde, gestellt und eine offene Frage, ob es technische Schwierigkeiten gab. Abschließend wurden die Teilnehmer\*innen in einer offenen Frage gebeten, das vermutete Ziel der Studie anzugeben.

### ***Pretest***

Im Vorfeld der Studie wurde ein Pretest durchgeführt um die Manipulation der Servicequalität zu testen. Der Pretest dauerte etwa 20 Minuten und wurde auf Englisch erhoben. Die Teilnehmer\*innen konnten einen von fünf Onlineshop-Gutscheinen gewinnen, der Wert richtete sich nach dem Ergebnis einer ausgelosten Runde des Steuerexperiments und betrug bis zu 25€. Die Steuerehrlichkeit wurde auf einer siebenstufige Likertskala als Intention, die Steuern ehrlich zu bezahlen („How likely would you be to pay your tax completely honestly?“), erhoben. Die erhobene Stichprobe besteht aus 83 Teilnehmer\*innen ( $n = 83$ ) von denen 24 männlich und 59 weiblich sind. Das durchschnittliche Alter beträgt 24.9 Jahre ( $SD = 5.28$ ). Die Teilnehmer\*innen wurden randomisiert entweder der hohen oder der niedrigen Serviceorientierung-Bedingung zugeteilt. Sie durchliefen 18 Steuerrunden, danach fand ein Manipulation-Check statt und am Schluss wurden vier Fragebögen und die soziodemographischen Daten erhoben. Die Teilnehmer\*innen der niedrigen SQ-Bedingung gaben eine signifikant niedrigere Zustimmung bei der Aussage an, dass die Steuerbehörde alles tue, um den Steuerzahler\*innen zu helfen ( $r = -.481, p < .01$ ). Auch den Aussagen, dass die Steuerbehörde die Steuerzahler\*innen respektvoll behandle ( $r (83) = -.487, p < .01$ ) und den Prozess so einfach wie möglich gestalte ( $r (83) = -.526, p < .01$ ), wurde bei niedriger SQ signifikant weniger zugestimmt. Darüber hinaus wurde auch das Interesse an der Meinung der Steuerzahler\*innen signifikant niedriger eingeschätzt ( $r (83) = -.196, p = .037$ ). Lediglich die Bewertung der Fairness des Prozesses unterscheidet sich nicht signifikant ( $r (83) = -.173, p = .057$ ) zwischen den Bedingungen. Aufgrund der Ergebnisse des Pretest kann angenommen werden, dass die gewählte Manipulation die wahrgenommene Serviceorientierung signifikant beeinflusst und diese in der Bedingung der niedrigen Servicequalität signifikant niedriger ist als in der Bedingung der hohen Servicequalität.

## Ergebnisse

Die Ergebnisse werden in drei Teilen vorgestellt, zu Beginn werden die Ergebnisse der Outcome-Ebene dargestellt, dann die Ergebnisse der Prozessebene und anschließend die explorative Auswertung. Aufgrund der Messwiederholungen an den Teilnehmer\*innen wurde ein Random Intercept in alle Modelle mit aufgenommen.

### Outcome-Ergebnisse

Die Teilnehmer\*innen der hohen SQ haben die Service Orientierung signifikant höher wahrgenommen ( $M = 3.69, SD = 0,75, p > .001$ ), als die Teilnehmer\*innen der niedrigen SQ ( $M = 2.51, SD = 0.87$ ). Für H1 wurde erwartet, dass die relative Steuerehrlichkeit der Teilnehmer\*innen der hohen SQ höher ist, als die relative Steuerehrlichkeit der niedrigen SQ. Ich habe eine *Lineare Mixed Effects Regression* durchgeführt, mit der relativen Steuerehrlichkeit als abhängiger Variable und der Bedingung der SQ als unabhängiger Variable. Die Bedingung der Servicequalität hat keinen signifikanten Zusammenhang mit der relativen Steuerehrlichkeit ( $B = -0.04, 95\% CI [-0.10, 0.02], p = .223$ ).

H2 postuliert, dass es in der Bedingung der niedrigen SQ bei einer Veränderung der Prüfwahrscheinlichkeit oder der Höhe der Strafe eine höhere Varianz in der Steuerehrlichkeit gibt, als in der Bedingung der hohen SQ. Es wurde eine Lineare Mixed Effects Regression durchgeführt, mit der relativen Steuerehrlichkeit als abhängiger Variable und der Interaktion zwischen den Strafparametern und der Bedingung der SQ als unabhängiger Variable. Es zeigt sich ein signifikanter Zusammenhang ( $B = 0.05, p = .002$ ) zwischen der Interaktion der 10-prozentigen Prüfwahrscheinlichkeit und der Bedingung der SQ und der relativen Steuerehrlichkeit. In der Bedingung der niedrigen SQ wirkt der Anstieg von der 5- auf die 10-prozentige und auf die 15-prozentige Prüfwahrscheinlichkeit stärker auf die Steuerehrlichkeit, als in der Bedingung der hohen SQ (s. Tab. 1). Der geschätzte Einfluss der Gruppenzugehörigkeit ist allerdings sehr klein. Es zeigte sich keine signifikante Interaktion zwischen Strafhöhe und der Bedingung der SQ auf die relative Steuerehrlichkeit (s. Tab. 1).

## Tabelle 1

*Strafparameter und Interaktion mit der Versuchsbedingung als Prädiktoren der Steuerehrlichkeit.  
Intercept für N = 485.*

Predictors	Estimates	95% CI	p
(Intercept)	0.40	0.36 – 0.45	<.001
Condition [SQ low]	-0.07	-0.13 – -0.01	.025
Audit probability [10%]	0.14	0.11 – 0.16	<.001
Audit probability [15%]	0.26	0.24 – 0.28	<.001
Fine rate [missing + 100%]	0.04	0.03 – 0.06	<.001
Condition [SQ low]*audit probability [15%]	0.05	0.02 – 0.08	.002
Condition [SQ low]*audit probability [15%]	0.04	0.01 – 0.07	.006
Condition [SQ low]*fine rate [missing + 100%]	0.01	-0.01 – 0.03	.419

8609 Observations

In Abbildung 3 ist die relative Steuerehrlichkeit dargestellt, geordnet nach Variation der Parameter. In Runde 1, 7 und 13 ist jeweils die Steuerehrlichkeit unter der geringsten Prüfwahrscheinlichkeit von 0.05 dargestellt und steigt nach zwei Runden erst auf 0.1, dann auf 0.15. Es ist erkennbar, dass mit Anstieg der Prüfwahrscheinlichkeit die Steuerehrlichkeit ansteigt, bis sie in Runde 7 und 13 wieder stark fällt.

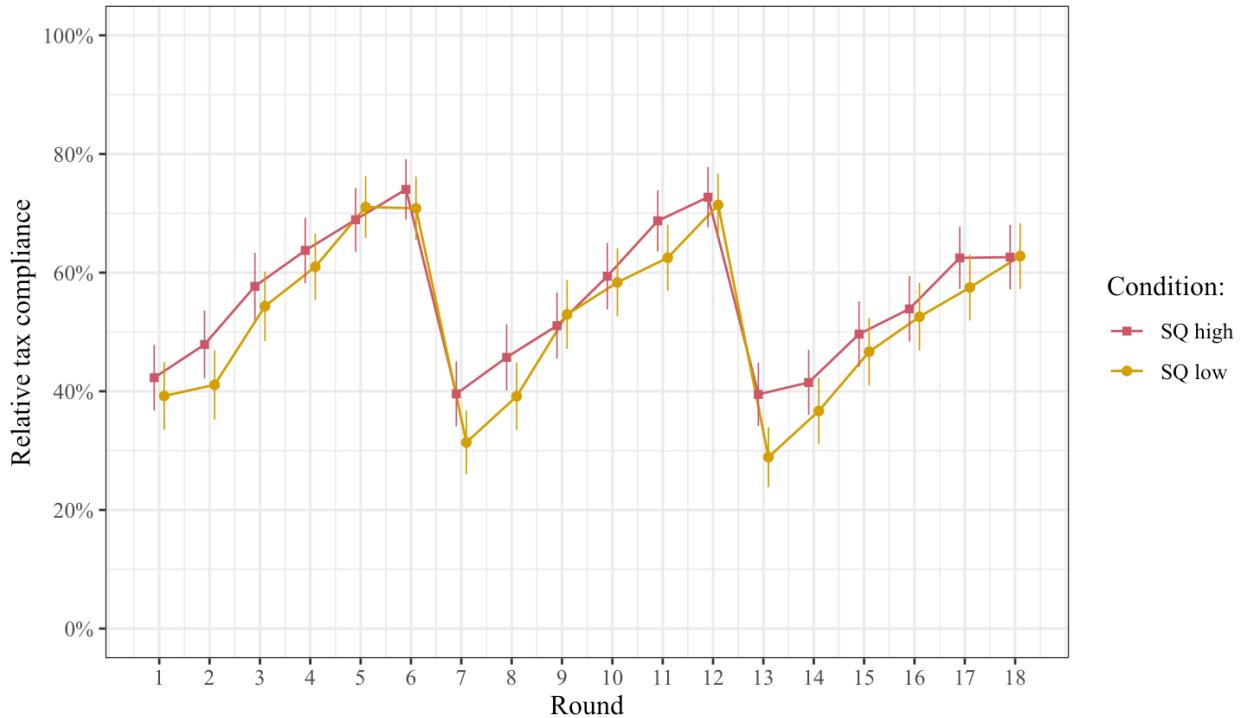
## Prozess-Ergebnisse

Für H3 wurde erwartet, dass aufgrund von höherer Aufmerksamkeit auf den Strafparametern, die Boxen der Prüfwahrscheinlichkeit und der Strafhöhe in der niedrigen SQ-Bedingung länger und häufiger geöffnet werden, als in der hohen SQ-Bedingung. Es wurde eine Lineare Mixed Effects Regression gerechnet, mit der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit und der Höhe der Strafe als abhängiger Variable und der Bedingung der SQ als unabhängiger Variable. Die Dauer der Boxöffnungen wurde logarithmisch transformiert, um eine Verteilung der Variablen zu erhalten, die einer Normalverteilung näher kommt (vgl. Schulte-Mecklenbeck et al., 2013). Unter Aufmerksamkeit auf den Strafparametern sind vier abhängige Variablen gefasst.

### Abbildung 3

Relative Steuerehrlichkeit nach Parametern. Legende der Parameter s. Tab. 2.

Die Runden wurden in der Erhebung für die Teilnehmer\*innen randomisiert dargestellt.



### Tabelle 2

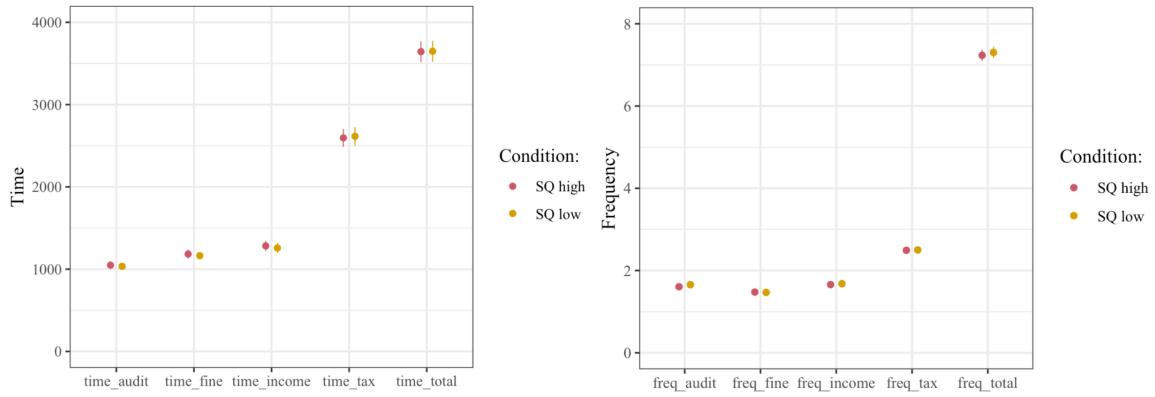
Parameter je Runde (ad. Abb. 3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
tax rate	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	
Audit probability	0.05	0.05	0.10	0.10	0.15	0.15	0.05	0.05	0.10	0.10	0.15	0.15	0.05	0.05	0.10	0.10	0.15	0.15
fine rate	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1

Es gibt weder einen signifikanten Zusammenhang zwischen Versuchsbedingung und Dauer der Boxöffnungen der Prüfwahrscheinlichkeit ( $B = 0.10$ , 95% CI [-0.15, 0.35],  $p = .422$ ), noch einen signifikanten Zusammenhang zwischen Versuchsbedingung und Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit ( $B = 0.05$ , 95% CI [-0.07, 0.17],  $p = .397$ ). Auch die Dauer der Boxöffnungen der Strafhöhe ( $B = 0.08$ , 95% CI [-0.25, 0.42],  $p = .632$ ) und die Häufigkeit der Boxöffnungen der Strafhöhe ( $B = -0.00$ , 95% CI [-0.13, 0.13],  $p = .951$ ), weisen keinen signifikanten Zusammenhang mit der Versuchsbedingung auf. Die durchschnittliche Öffnungsduer und -häufigkeit der Boxen mit den korrespondierenden Parametern ist in Abbildung 4 dargestellt.

#### Abbildung 4

Durchschnittliche Dauer (in ms) und Häufigkeit der Öffnungen der Parameter je Versuchsbedingung.  $N = 485$ .



H4 postuliert, dass in der niedrigen SQ-Bedingung mehr Übergänge zwischen der Prüfwahrscheinlichkeit und der Strafhöhe stattfinden, als in der hohen SQ-Bedingung. Es wurde eine Lineare Mixed Effects Regression gerechnet, mit der Anzahl der Übergänge zwischen den Boxen der Prüfwahrscheinlichkeit und den Boxen der Strafhöhe als abhängiger Variable und der Bedingung der SQ als unabhängiger Variable. Es gibt keinen signifikanten Unterschied der Anzahl der Übergänge zwischen den Teilnehmer\*innen der beiden Bedingungen ( $B = 0.03$ , 95% CI [-0.08, 0.15],  $p = .584$ ).

H5 postuliert, dass die Teilnehmer\*innen der niedrigen SQ-Bedingung längere Entscheidungszeiten pro Runde haben, als die Teilnehmer\*innen der hohen SQ-Bedingung. Es wurde eine Lineare Mixed Effects Regression gerechnet, mit der Entscheidungszeit je Runde als abhängiger Variable und der Bedingung der SQ als unabhängiger Variable. Es gibt keinen signifikanten Unterschied der Entscheidungszeiten zwischen den Teilnehmer\*innen der beiden Bedingungen ( $B = 0.04$ , 95% CI [-0.08, 0.15],  $p = .529$ ).

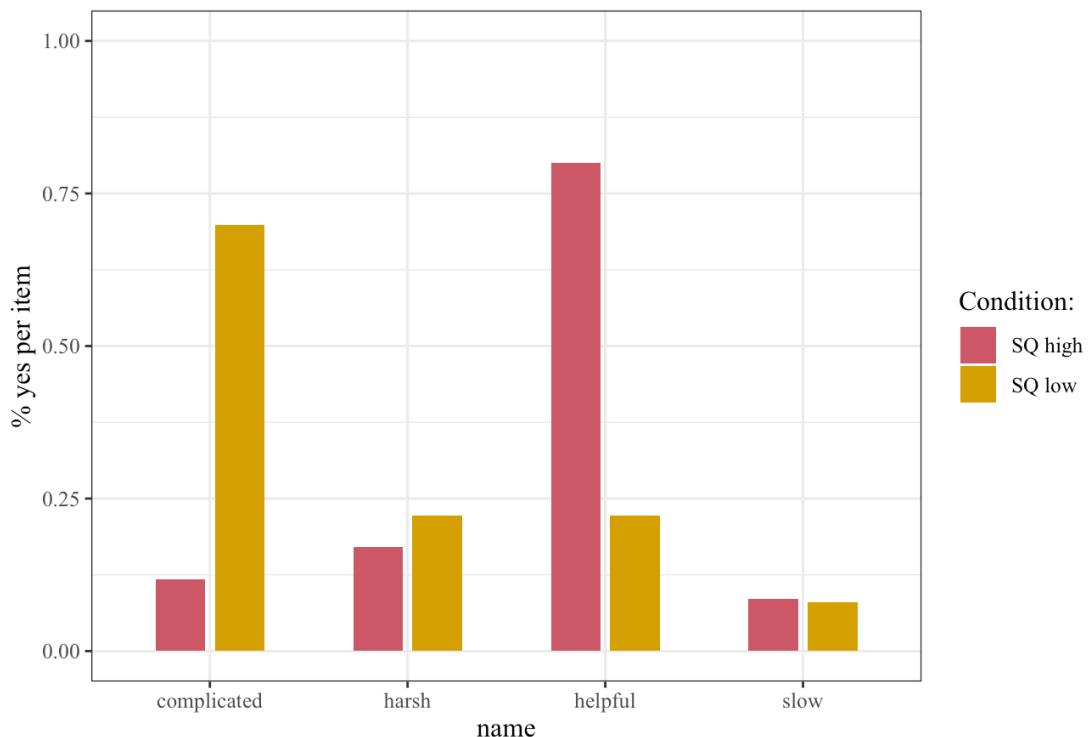
## Explorative Auswertung

### *Manipulation der Servicequalität*

Mittels einer Multiple-Choice-Frage wurde erhoben, ob die Teilnehmer\*innen sich an die Vorstellung der Steuerbehörde erinnern. Sowohl die Bedingung der niedrigen SQ (69.89% vs. 11.85%,  $p < .001$ ), als auch die Bedingung der hohen SQ (80.22% vs. 22.15%,  $p < .001$ ) zeigen eine deutlich höhere durchschnittliche Zustimmung zu den passenden Items (s. Abb. 5).

**Abbildung 5**

*Prozentualer Anteil der Zustimmung zu den vier Items des Manipulation-Check je Versuchsbedingung. N = 485*



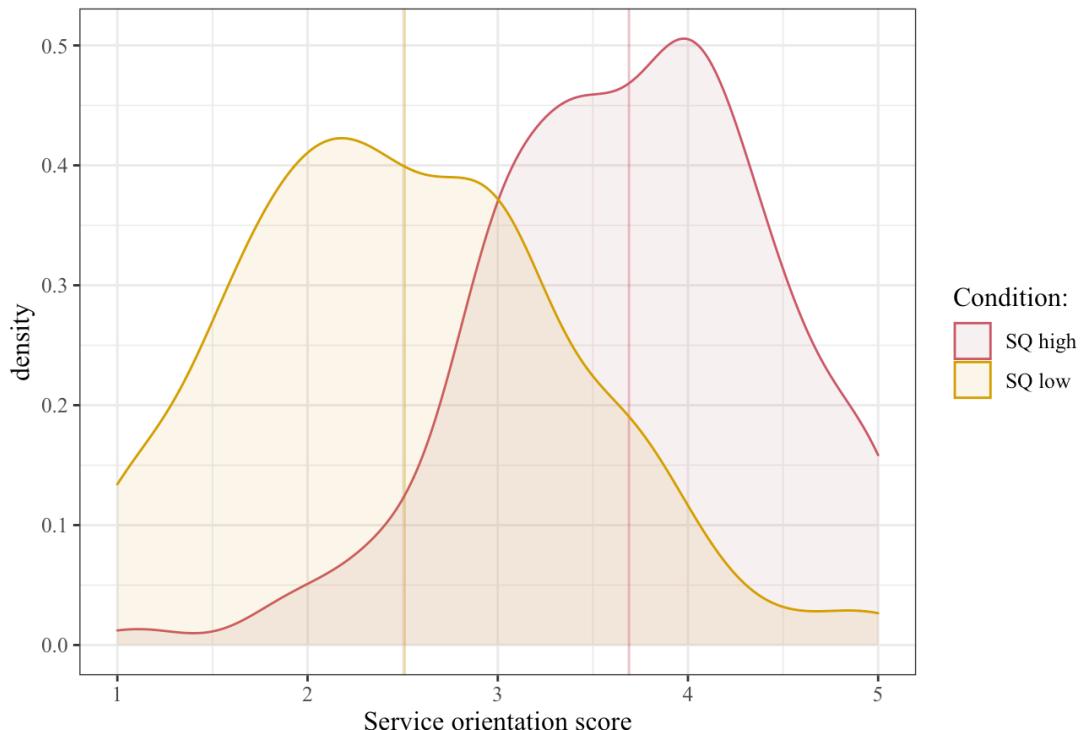
Auch die vier Fragen zur Erhebung des wahrgenommenen Aufwands der Teilnehmer\*innen (*complicated, annoying, difficult, concentrate*) wurden in der Bedingung der niedrigen SQ mit  $M = 3.05$  ( $SD = 0.88$ ) signifikant höher angegeben, als in der Bedingung der hohen SQ mit  $M = 2.48$  ( $SD = 0.97$ ,  $p < .001$ ).

Der Service Orientation Score (Gangl et al., 2013) ist in der Bedingung der hohen SQ mit  $M = 3.69$  ( $SD = 0.75$ ) ebenfalls signifikant höher ( $p < .001$ ), als in der Bedingung der niedrigen SQ mit  $M = 2.51$  ( $SD = 0.87$ ) (s. Abb. 6).

**Abbildung 6**

*Perceived Service Orientation Score je Versuchsbedingung.*

$N = 485$



### ***Selbstangegebene Wichtigkeit der Parameter***

Es wurden Lineare Mixed Effects Regressionen durchgeführt, um zu untersuchen, ob die selbstangegebene Wichtigkeit der vier Parameter für die Entscheidung (Schieberegler von *not important at all* bis *very important*, kodiert von 1-101) die Dauer bzw. die Häufigkeit der Boxöffnungen vorhersagt.

Zwischen Wichtigkeit und Dauer der Boxöffnungen der Prüfwahrscheinlichkeit besteht ein signifikanter positiver Zusammenhang ( $B = 2.60$ ,  $p = .001$ ), geben Teilnehmer\*innen die Wichtigkeit der Prüfwahrscheinlichkeit um 10 Punkte höher an, beträgt die geschätzte Dauer der Boxöffnung 26 ms mehr. Auch der Zusammenhang

zwischen Wichtigkeit und Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit ist signifikant ( $B = 0.004, p < .001$ ), jedoch ist der Schätzer sehr klein und bedeutet somit bei einer um 10 Punkte höher angegebenen Wichtigkeit lediglich 0.04 Boxöffnungen mehr.

Der Zusammenhang zwischen der Wichtigkeit und der Dauer der Boxöffnungen der Strafhöhe ( $B = 7.21, p < .001$ ), sowie der Zusammenhang zwischen der Wichtigkeit und der Häufigkeit der Boxöffnungen der Strafhöhe ( $B = 0.01, p < .001$ ) sind signifikant. Teilnehmer\*innen, die eine höhere Wichtigkeit der Strafhöhe angeben, öffnen auch diese Box länger und häufiger.

Auch der Zusammenhang zwischen Wichtigkeit und Dauer der Boxöffnungen der Steuerrate/-höhe ( $B = 9.63, p = .001$ ) und der Zusammenhang zwischen Wichtigkeit und Häufigkeit der Boxöffnungen der Steuerrate/-höhe ( $B = 0.004, p = .003$ ) sind signifikant. Teilnehmer\*innen, die eine höhere Wichtigkeit der Steuerrate/-höhe angeben, öffnen auch diese Box länger und häufiger.

Lediglich zwischen der Wichtigkeit und der Dauer ( $B = 0.73, p = .536$ ) bzw. der Häufigkeit ( $B = 0.00, p = .20$ ) der Boxöffnungen des Einkommens bestehen keine signifikanten Zusammenhänge.

### ***Risikoneigung***

Nach dem Steuerexperiment gaben die Teilnehmer\*innen beider Bedingungen im postexperimentellen Fragebogen an, ob sie bereit sind Risiken einzugehen, oder ob sie dies vermeiden (Likert-Skala von  $1 = \text{Definitely avoiding to take risks}$  bis  $9 = \text{Fully prepared to take risks}$ ), erhoben nach Dohmen et al. (2011). Die Teilnehmer\*innen haben durchschnittlich ihre Risikoneigung mit  $M = 4.95 (SD = 2.49)$  bewertet. Die Teilnehmer\*innen der niedrigen SQ haben ihre Risikoneigung mit  $M = 5.09$  signifikant höher angegeben, als die Teilnehmer der hohen SQ mit  $M = 4.82 (t(20.56) = 35.84, p < .001)$ . Es zeigt sich jedoch kein signifikanter Interaktionseffekt zwischen Risikoneigung und Versuchsbedingung auf die Steuerehrlichkeit oder auf die Dauer oder Häufigkeit der Boxöffnungen der Strafparameter (s. Anhang B3).

### ***Erzwungene vs. Freiwillige Steuerehrlichkeit***

Die Teilnehmer\*innen der beiden Bedingungen zeigen zwar keinen signifikanten Unterschied in der erzwungenen Steuerehrlichkeit (SE), jedoch einen signifikanten Unterschied in der freiwilligen SE (s. Tab. 3). Darüber hinaus korreliert die freiwillige SE signifikant mit den Öffnungszeiten und der Öffnungshäufigkeit beider Strafparameter (s. Anhang B5). Es zeigt sich jedoch kein signifikanter Interaktionseffekt zwischen freiwilliger SE und Versuchsbedingung auf die Steuerehrlichkeit oder auf die Dauer oder Häufigkeit der Boxöffnungen der Strafparameter (s. Anhang B5).

### **Tabelle 3**

*Erzwungene und freiwillige Steuerehrlichkeit im Vergleich der beiden Versuchsbedingungen. Ausprägungen von 1 bis 5.*

	Erzwungene SE	Freiwillige SE
Hohe SQ	$M = 3.21 (SD = 1.01)$	$M = 3.46 (SD = 1.17)$
Niedrige SQ	$M = 3.31 (SD = 1.08)$	$M = 3.24 (SD = 1.21)$
	$p = .323$	$p = .042$

### ***Aufmerksamkeits-Check***

Es wurden das Verständnis der Steuerparameter und die Aufmerksamkeit während der Erhebung mit jeweils vier Items erhoben. Die beiden Bedingungen unterscheiden sich nicht signifikant im Verständnis der Parameter, jedoch in der Aufmerksamkeit während der Erhebung (s. Tab. 4). Es zeigt sich jedoch kein signifikanter Interaktionseffekt zwischen der Aufmerksamkeit und der Versuchsbedingung auf die Steuerehrlichkeit oder die Betrachtungsdauer (s. Anhang B4).

**Tabelle 4**

*Verständnis der Parameter und Aufmerksamkeit während der Erhebung im Vergleich der beiden Bedingungen.*

	Verständnis (0 bis 4)	Aufmerksamkeit (1 bis 5)
Hohe SQ	$M = 3.16 (SD = 0.89)$	$M = 3.79 (SD = 0.39)$
Niedrige SQ	$M = 3.24 (SD = 0.87)$	$M = 3.89 (SD = 0.36)$
	$p = .318$	$p = .003$

### ***Sozio-demografische Variablen***

Das Alter der Teilnehmer\*innen sagt sowohl die relative Steuerehrlichkeit ( $B = 0.00, p = .046$ ) als auch die Häufigkeiten der Boxöffnungen der Prüfwahrscheinlichkeit ( $B = -0.01, p < .001$ ) und der Strafhöhe ( $B = -0.01, p < .001$ ) signifikant vorher. Es gibt allerdings keine signifikanten Interaktionseffekte zwischen Alter und Versuchsbedingung (s. Anhang B1). Es zeigt sich weder zwischen Geschlecht und relativer Steuerehrlichkeit noch zwischen Geschlecht und der Dauer oder der Häufigkeit der Boxöffnungen der beiden Strafparameter ein signifikanter Zusammenhang, und darüber hinaus zeigen sich auch keine signifikanten Interaktionseffekte zwischen Geschlecht und Versuchsbedingung (s. Anhang B2).

### **Diskussion**

Das Ziel der vorliegenden Arbeit bestand darin, eine Einsicht in die Prozesse zu ermöglichen, die der messbaren Variable der Steuerehrlichkeit zugrunde liegen. Die Steuerehrlichkeit, das Ergebnis der Steuerentscheidung, ist ein umfassend untersuchtes Feld, jedoch mit widersprüchlichen Ergebnissen. Die Analyse der Prozesse, die der Steuerentscheidung voraus gehen, ist ein vielversprechender Ansatz, um Einblicke zu erhalten, wie Informationen die Steuerentscheidung beeinflussen. Auf Outcome-Ebene habe ich angenommen, dass bei hoher SQ die Steuerehrlichkeit höher sei und die Varianz der Steuerehrlichkeit, bei Variation der Strafparameter, niedriger. Entgegen meiner Erwartung zeigten sich weder signifikante Unterschiede zwischen den beiden Bedingungen in der relativen Steuerehrlichkeit, noch hatten Veränderungen der Parameter signifikant unterschiedliche Auswirkungen auf die relative Steuerehrlichkeit. Auch auf

Prozess-Ebene zeigten sich entgegen meiner Erwartungen keine Unterschiede in der Dauer oder Häufigkeit der Betrachtung der Strafparameter, Übergängen zwischen diesen oder der Entscheidungszeit je Runde zwischen der hohen und niedrigen SQ Bedingung.

Somit stehen meine Resultate im Gegensatz zu bisherigen Belegen für die positiven Zusammenhänge zwischen Servicequalität und Steuerehrlichkeit (Feld & Frey, 2015, Alm et al., 2010, Alm & Torgler, 2011, Gangl et al., 2013). Allerdings wurden diese Zusammenhänge zumeist nicht in experimentellen Erhebungen, sondern im Bezug auf reale Steuerbehörden gefunden. Zum einen scheint hier ein entscheidender Unterschied in der Funktion der Steuergelder zu liegen, da sich die Steuerehrlichkeit von Teilnehmer\*innen z.B. leichter durch soziale Normen (Wenzel, 2004) oder Macht der Behörden (Alm et al., 2012) beeinflussen lassen, wenn die gezahlten Steuern dem Gemeinwohl gewidmet sind. Zum anderen könnte möglicherweise die Wahrnehmung der Steuerbehörde im experimentellen Setting zu sehr von der Wahrnehmung der realen Steuerbehörde beeinflusst sein. Dies würde erklären, warum sich die erwarteten Effekte zwar zwischen relativer Steuerehrlichkeit bzw. Öffnungszeiten/-dauer der Strafparameter und der Art der Steuerehrlichkeit (freiwillig vs. erzwungen) zeigen, aber nicht zwischen den beiden Bedingungen.

Die Manipulation der Servicequalität hat zwar funktioniert, die Teilnehmer\*innen der niedrigen Bedingung nahmen die Steuerbehörde signifikant weniger serviceorientiert wahr, allerdings liegt die wahrgenommene Service Orientierung auch in der niedrigen Bedingung immer noch im mittleren Bereich. Außerdem erinnerten sich die Teilnehmer\*innen gut an die jeweilige Vorstellung der Steuerbehörde und der Prozess der Steuererhebung wird in der Bedingung der niedrigen SQ signifikant aufwendiger wahrgenommen. Insgesamt scheint also das Ziel einer schlechteren Servicequalität in der niedrigen Bedingung erreicht worden zu sein. Darüber hinaus wurde neben der Dauer und der Häufigkeit der Boxöffnungen auch die eigene Einschätzung der Teilnehmer\*innen erhoben, wie wichtig die einzelnen Parameter für ihre Entscheidung waren. Bis auf das Einkommen zeigen sich bei allen Parametern signifikante Zusammenhänge zwischen der angegeben Wichtigkeit und der Dauer und Häufigkeit der Boxöffnungen. MouselabWEB scheint gut geeignet, die Aufmerksamkeit durch die Öffnungszeiten und -dauern zu

operationalisieren, allerdings wurden weder die Hypothesen bezüglich des Outcomes noch die Hypothesen bezüglich des Prozesses der Boxöffnungen bestätigt.

Die Art der Steuerehrlichkeit (freiwillig vs. erzwungen; Kirchler & Wahl, 2010) zeigt hingegen sowohl signifikante Zusammenhänge mit der Öffnungsduer und Häufigkeit beider Strafparameter als auch mit der relativen Steuerehrlichkeit. Bei höherer freiwilliger Steuerehrlichkeit, werden beide Strafparameter kürzer und seltener angeschaut und es wird eine höhere relative Steuerehrlichkeit erreicht. Dies stützt die Annahmen des Slippery Slope Modells (Kirchler et al., 2008), laut dem freiwillige Steuerehrlichkeit durch Vertrauen in die Behörde entsteht und nicht aus Furcht vor Strafe. Allerdings unterscheidet sich die Art der Steuerehrlichkeit nicht signifikant zwischen den Versuchsbedingungen. Laut dem Slippery Slope Modell, sollte die Dimension des Vertrauens in die Steuerbehörde kaum Einfluss haben, wenn die Macht der Steuerbehörde als sehr stark wahrgenommen wird. Möglicherweise wurde die Steuerbehörde in der durchgeführten Erhebung so mächtig wahrgenommen, dass aus diesem Grund die Variation der Servicequalität keinen Einfluss mehr zeigt. Durch das Design am Computer, ist anzunehmen, dass bei einer Prüfung eine Steuerhinterziehung mit Sicherheit entdeckt würde und auch die Strafe bezahlt werden muss, da sie automatisch bei der Auszahlung einbehalten wird. Dies könnte dazu führen, dass die Steuerbehörde als so mächtig wahrgenommen wird, dass die Unterschiede in der wahrgenommenen Serviceorientierung nicht stark genug sind, darüber hinaus noch einen Effekt auf die Steuerehrlichkeit zu nehmen. In zukünftigen Studien könnte zusätzlich noch die subjektiv wahrgenommene Prüfwahrscheinlichkeit (Kirchler et al., 2008; Alm, 2012) nach jeder Runde erhoben werden, da Personen darüber hinaus dazu neigen, niedrige Prüfwahrscheinlichkeiten zu überschätzen (Alm, 2012).

Darüber hinaus könnte in zukünftigen Projekten auch die Servicequalität noch stärker beeinträchtigt werden, z.B. durch einen noch umständlicheren Prozess der Umkodierung. Zusätzlich könnte auch die wahrgenommene Fairness stärker beeinflusst werden, z.B. indem den Teilnehmer\*innen der niedrigen Bedingung mitgeteilt wird, dass manche Teilnehmer\*innen die Umkodierung selber ausführen müssen und dies für andere Teilnehmer\*innen automatisiert geschieht. Darüber hinaus könnte man zusätzlich die wahrgenommene Macht der Steuerbehörde erheben.

## Fazit

Die Teilnehmer\*innen der niedrigen SQ haben die Steuerbehörde zwar signifikant weniger serviceorientiert wahrgenommen, jedoch liegt die wahrgenommene Serviceorientierung in der Versuchsbedingung der niedrigen SQ immer noch im mittleren Bereich, in weiteren Studien könnte die Fragestellung mittels einer stärkeren Manipulation der SQ untersucht werden. Darüber hinaus zeigen die explorativen Ergebnisse, dass die Einstellung zu Steuern bzw. somit die Wahrnehmung der Steuerbehörde einen Einfluss auf die Aufmerksamkeit auf den Strafparametern haben könnten. Dies zeigt, wie vielversprechend Prozessanalysen als Forschungsinstrument sein können, um die Grundannahmen von Steuerehrlichkeitsmodellen zu prüfen. Diese sollten, gerade in einem Feld mit so widersprüchlichen Ergebnissen, verbreiteter Anwendung finden.

## Danksagung

An erster Stelle möchte ich mich herzlich bei ██████████ bedanken, für seine Zeit und Mühe als Korrekturleser und dafür, dass er mich richtungsweisend und mit viel Engagement während meiner Arbeit begleitet hat und die Überlassung des Themas. Seine konstruktive Kritik und allzeit offenes Ohr haben mir sehr geholfen!

Des Weiteren möchte ich mich bei ██████████ edanken, für viele wertvolle Ratschläge und die kontinuierliche Betreuung über den ganzen Prozess hinweg. Auch dem Studienpräses der Universität Wien gilt ein besonderer Dank für die finanzielle Unterstützung!

## Referenzen

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of public economics, 1(3-4)*, 323-338. doi: 10.1016/0047-2727(72)90010-2

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance, 19(1)*, 54-77. doi: 10.1007/s10797-011-9171-2

Alm, J., Cherry, T., Jones, M., & McKee, M. (2010). Taxpayer information assistance services and tax compliance behavior. *Journal of Public Economics, 31*, 577-586. doi:10.1016/j.joep.2010.03.018

Alm, J., McClelland, G.H., & Schulze, W.D. (1992). Why do people pay taxes? *Journal of Public Economics, 48(1)*, 21-38. doi: 10.1016/0047-2727(92)90040-M

Alm, J., Kirchler, E., & Muehlbacher, S. (2012): Combining Psychology and Economics in the analysis of compliance: From enforcement to cooperation. *Economic Analysis & Policy, 42 (2)*, 133–151.

Batrancea, L., Nichita, A., Olsen, J., Kogler, C., Kirchler, E., Hoelzl, E.,... Zukauskasbo,S. (2019). Trust and power as determinants of tax compliance across 44 nations. *Journal of Economic Psychology, 74*. doi: 10.1016/j.joep.2019.102191

Bouckaert, G., & van de Walle, S. (2003). Comparing Measures of Citizen Trust and User Satisfaction as Indicators of “Good Governance”: Difficulties in Linking Trust and Satisfaction Indicators. *International Review of Administrative Science 69*, 329-343. doi: 10.1177/0020852303693003

Braithwaite, V. (2003a). Dancing with tax authorities: Motivational postures and non-compliant actions. *Taxing democracy*, 15-39. Denemark, D. (2011). Trust, Efficacy

and Opposition to Anti-terrorism Police Powers: Australia in Comparative Perspective. European Consortium for Political Research General Conference.

Braithwaite, V. (2003b), A New Approach to Tax Compliance, in: Braithwaite, V. (Ed.), *Taxing Democracy. Understanding Tax Avoidance and Tax Evasion*, Ashgate, Aldershot, UK, 1–11.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69 (5), 1193-1235. doi: 10.1111/1468-0262.00239

van Dijke, M., Gobena, L.B.&Verboon, P. (2019). Make Me Want to Pay. A Three-Way Interaction Between Procedural Justice, Distributive Justice, and Power on Voluntary Tax Compliance. *Frontiers in Psychology*, 10, 16-32. doi: 10.3389/fpsyg.2019.01632

Farrar, J. (2015). An empirical analysis of taxpayers fairness preferences from Canada's Taxpayer Bill of rights. *Journal of Accounting and Taxation*, 7, 71–79. doi: 10.5897/jat2015.0183

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. doi: 10.3758/BF03193146

Feld, L.P.,& Frey, B.S. (2005). Feld, L., & Frey, B. (2005). Illegal, immoral, fattening or what?: How deterrence and responsive regulation shape tax morale. In C. Bajada & F. Schneider (Hrsg.). *Size. causes, and consequences of underground economy. An international perspective* (S. 16–37). Aldershot: Ashgate.

Fischer, C.M., Wartick, M., & Mark, M. (1992). Detection Probability and Taxpayer Compliance: A Review of the Literature. *Journal of Accounting Literature*, 11, 1-46. doi: 10.12691/jfa-2-1-1

Gangl, K., Muehlbacher, S., de Groot, M., Goslinga, S., Hofman, E., Kogler, C., ... Kirchler, E. (2013). "How can I help you?" Perceived Service Orientation of Tax Authorities and Tax Compliance. *Public Finance Analysis*, 69(4), 487-510. doi: 10.1628/001522113X675683

Hartner, M., Rechberger, S., Kirchler, E., and Schabmann, A. (2008). Procedural fairness and tax compliance. *Econ. Anal. Policy* 38, 137-152. doi: 10.1016/s0313-5926(08)50010- 5

Kirchler, E. (2007). The economic psychology of tax behavior. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511628238

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, 29(2), 210-225. doi: 10.1016/j.joep.2007.05.004

Kirchler, E., & Wahl, I. (2010). Tax compliance inventory TAX-I: Designing an inventory for surveys of tax compliance. *Journal of Economic Psychology*, 31, 331-346. doi: 10.1016/j.joep.2010.01.002

Kogler, C., Olsen, J., Müller, M., & Kirchler, E. (2022). Information processing in tax decisions: a MouselabWEB study on the Allingham and Sandro model of income tax evasion. *Journal of Behavioral Decision Making*, 2022, 1-15. doi: 10.1002/bdm.2272

Mittone, L. (2005). Dynamic behavior in tax evasion: An experimental approach. *The Journal of Socio-Economics*, 35, 813-835. doi: 10.1016/j.soec.2005.11.065

Murphy, K. (2003). The importance of procedural justice in securing compliance: A study of tax scheme investors. *Australian Journal Of Social Issues*, 38(3), 365-390.

Nielsen, M., Haun, D., Kartner, J., & Legare, C. H. (2007). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 102, 31-38. doi: 10.1016/j.jecp.2007.04.017.

Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, 71, 242-251. doi: 10.1016/j.appet.2013.08.008

Shou, Y., & Olney, J. (2020). Assessing a domain-specific risk-taking construct: A meta-analysis of reliability of the DOSPERT scale. *Judgement and Decision Making*, 15(1), 112-134

Statistik Austria (14.1.2022). Steuern und Sozialbeiträge in Österreich, Einnahmen des Staates und der EU [Tabelle]. Verfügbar unter [https://www.statistik.at/web\\_de/statistiken/wirtschaft/oeffentliche\\_finanzen\\_und\\_steuern/oeffentliche\\_finanzen/steuereinnahmen/019086.html](https://www.statistik.at/web_de/statistiken/wirtschaft/oeffentliche_finanzen_und_steuern/oeffentliche_finanzen/steuereinnahmen/019086.html)

Tyler, T. (2006). Psychological perspectives on legitimacy and legitimization. *Annual Review of Psychology*, 57, 375–400. doi: 10.1146/annurev.psych.57.102904.190038

Weber, E.U., Blais, A.-R., & Betz, N. (2002). DOSPERT. Domain-Specific Risk Taking. *Journal of Behavioral Decision Making*, 15, 263-290.

Wenzel, M. (2004): An analysis of norm processes in tax compliance. *Journal of Economic Psychology*, 25 (2), 213–228.

Willemse, M. C., & Johnson, E. J. (2019). Observing Cognition with MouselabWEB. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Hrsg.), *A handbook of process tracing methods* (S. 76–95). Routledge.

## Anhang A

### Link zu dem Material der Erhebung

Eine vollständige Darstellung der durchgeführten Erhebung kann unter folgender Adresse abgerufen werden: <https://osf.io/rxzcu/>

## Anhang B

### B1 Alter und Versuchsbedingung

#### Tabelle B1.1

*Alter und Versuchsbedingung als Prädiktoren der relativen Steuerehrlichkeit. Intercept für N = 482.*

Predictors	Estimates	p
(Intercept)	0.44	< 0.001***
condition [SQ low]	0.01	0.902
age prolific	0.00	0.046*
condition [SQ low]*age prolific	-0.00	0.543
8555 Observations		

#### Tabelle B1.2

*Alter und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log). Intercept für N = 482.*

Predictors	Dauer der Boxöffnungen der Prüfwahrscheinlichkeit (log)		Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log)	
	Estimates	p	Estimates	p
(Intercept)	6.57	< 0.001***	2.00	< 0.001***
condition [SQ low]	-0.61	0.101	-0.17	0.320
age prolific	-0.01	0.190	-0.01	< 0.001***
condition [SQ low] * age prolific	0.02	0.051	0.01	0.163
8071 Observations				

**Tabelle B1.4**

*Alter und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Strafhöhe (log). Intercept für N = 482.*

Predictors	Dauer der Boxöffnungen der Strafhöhe (log)		Häufigkeit der Boxöffnungen der Strafhöhe (log)	
	Estimates	p	Estimates	p
(Intercept)	5.90	< 0.001***	1.87	< 0.001***
condition [SQ low]	-0.61	0.219	-0.23	0.232
age prolific	-0.00	0.780	-0.01	0.001**
condition [SQ low] * age prolific	0.02	0.125	0.01	0.198
8071 Observations				

## B2 Geschlecht und Versuchsbedingung

**Tabelle B2.1**

*Geschlecht und Versuchsbedingung als Prädiktoren der Steuerehrlichkeit. Intercept für N = 485.*

Predictors	Estimates	p
(Intercept)	0.81	< 0.001
condition [SQ low]	-0.05	0.237
sex prolific [Female]	-0.22	0.338
condition [SQ low]*sex prolific	0.03	668
8609 Observations		

**Tabelle B2.2**

*Geschlecht und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log). Intercept für N = 485.*

Predictors	Dauer der Boxöffnungen der Prüfwahrscheinlichkeit (log)		Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log)	
	Estimates	p	Estimates	p
(Intercept)	6.91	< 0.001***	1.68	< 0.001***
condition [SQ low]	0.36	0.064	0.20	0.033*
sex prolific [Female]	-0.60	0.550	-0.10	0.838
condition [SQ low] * sex prolific [Female]	-0.44	0.090	-0.24	0.058
8115 Observations				

**Tabelle B2.4**

*Geschlecht und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Strafhöhe (log). Intercept für N = 485.*

Predictors	Dauer der Boxöffnungen der Strafhöhe (log)		Häufigkeit der Boxöffnungen der Strafhöhe (log)	
	Estimates	p	Estimates	p
(Intercept)	7.02	< 0.001***	1.59	0.002**
condition [SQ low]	0.30	0.259	0.08	0.445
sex prolific [Female]	-1.22	0.373	-0.20	0.712
condition [SQ low] * sex prolific [Female]	-0.35	0.322	-0.12	0.356
8115 Observations				

### B3 Risikoneigung

Erhoben nach Gangl et al., 2013.

#### Tabelle B3.1

*Risikoneigung und Versuchsbedingung als Prädiktoren der relativen Steuerehrlichkeit. Intercept für N = 485.*

<i>Predictors</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	0.75	< 0.001***
condition [SQ low]	-0.01	0.893
risk attitude	-0.04	< 0.001***
condition [SQ low] * risk attitude	-0.00	0.765
8609 Observations		

#### Tabelle B3.2

*Risikoneigung und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log). Intercept für N = 485.*

<i>Predictors</i>	Dauer der Boxöffnungen der Prüfwahrscheinlichkeit (log)		Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log)	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	6.35	< 0.001***	1.55	< 0.001***
condition [SQ low]	-0.16	0.573	-0.00	0.996
risk attitude	-0.02	0.596	0.01	0.661
condition [SQ low] * risk attitude	0.05	0.305	0.01	0.687
8115 Observations				

**Tabelle B3.3**

*Risikoneigung und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Strafhöhe (log). Intercept für N = 485.*

Predictors	Dauer der Boxöffnungen der Strafhöhe (log)		Häufigkeit der Boxöffnungen der Strafhöhe (log)	
	Estimates	p	Estimates	p
(Intercept)	6.09	< 0.001***	1.40	< 0.001***
condition [SQ low]	-0.38	0.323	0.03	0.824
risk attitude	-0.06	0.241	0.01	0.552
condition [SQ low] * risk attitude	0.09	0.176	-0.01	0.768
8115 Observations				

#### **B4 Aufmerksamkeit während der Erhebung**

**Tabelle B4.1**

*Aufmerksamkeit und Versuchsbedingung als Prädiktoren der relativen Steuerehrlichkeit. Intercept für N = 485.*

Predictors	Estimates	p
(Intercept)	0.91	< 0.001
condition [SQ low]	-0.28	0.371
attention check score	-0.09	0.084
condition [SQ low] * attention check score	0.06	0.423
8115 Observations		

**Tabelle B4.2**

*Aufmerksamkeit und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log). Intercept für N = 485.*

Predictors	Dauer der Boxöffnungen der Prüfwahrscheinlichkeit (log)		Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log)	
	Estimates	p	Estimates	p
(Intercept)	3.80	< 0.001***	0.61	0.146
condition [SQ low]	1.56	0.232	0.79	0.204
attention check score	0.65	0.005	0.26	0.018*
condition [SQ low] * attention check score	-0.39	0.246	-0.20	0.222

8115 Observations

**Tabelle B4.3**

*Aufmerksamkeit und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Strafhöhe (log). Intercept für N = 485.*

Predictors	Dauer der Boxöffnungen der Strafhöhe (log)		Häufigkeit der Boxöffnungen der Strafhöhe (log)	
	Estimates	p	Estimates	p
(Intercept)	3.50	3	0.67	0.147
condition [SQ low]	3.16	0.074	1.16	0.092
attention check score	0.61	0.051	0.21	0.085
condition [SQ low] * attention check score	-0.81	0.078	-0.30	0.087

8115 Observations

## B5 Art der Steuerehrlichkeit

*Freiwillige vs. erzwungene SE erhoben nach Kirchler & Wahl, 2010.*

### Tabelle B5.1

*Art der SE und Versuchsbedingung als Prädiktoren der relativen Steuerehrlichkeit.  
Intercept für N = 485.*

<i>Predictors</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	-0.03	0.537
condition [SQ low]	-0.03	0.622
ssf voluntary	0.17	< 0.001***
condition [SQ low] * ssf voluntary	0.01	0.577

8609 Observations

### Tabelle B5.2

*Art der SE und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log). Intercept für N = 485.*

<i>Predictors</i>	Dauer der Boxöffnungen der Prüfwahrscheinlichkeit (log)		Häufigkeit der Boxöffnungen der Prüfwahrscheinlichkeit (log)	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	7.00	< 0.001***	2.09	< 0.001***
condition [SQ low]	0.16	0.671	-0.11	0.545
ssf voluntary	-0.22	0.004*	-0.15	< 0.001***
condition [SQ low] * ssf voluntary	-0.03	0.761	0.04	0.431

8115 Observations

**Tabelle B5.3**

*Art der SE und Versuchsbedingung als Prädiktoren der Dauer und der Häufigkeit der Boxöffnungen der Strafhöhe (log). Intercept für N = 485.*

<i>Predictors</i>	Dauer der Boxöffnungen der Strafhöhe (log)		Häufigkeit der Boxöffnungen der Strafhöhe (log)	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	6.80	< 0.001***	1.99	< 0.001***
condition [SQ low]	0.21	0.681	-0.18	0.344
ssf voluntary	-0.29	0.005*	-0.15	< 0.001***
condition [SQ low] * ssf voluntary	-0.06	0.683	0.05	0.408

---

8115 Observations