Journal of Economic Behavior and Organization 185 (2021) 250–286

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Dishonesty and risk-taking: Compliance decisions of individuals and groups ☆

Martin Fochmann[a,b,*], Nadja Fochmann[c], Martin G. Kocher[d,e,f], Nadja Müller[g]

[a] School of Business & Economics, Free University of Berlin, Thielallee 73, D-14195 Berlin, Germany
[b] Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany
[c] Faculty of Economics and Management, University of Hanover, Königsworther Platz 1, D-30167 Hanover, Germany
[d] Institute for Advanced Studies, Vienna, Austria
[e] Department of Economics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria
[f] University of Gothenburg, Gothenburg, Sweden
[g] Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany

ARTICLE INFO

Article history:
Received 17 February 2020
Revised 7 February 2021
Accepted 9 February 2021
Available online 23 March 2021

JEL classification:
C91
C92
D03
H26

Keywords:
Dishonesty
Lying
Compliance
Risk-taking
Group decisions
Communication
Norms
Experiment

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 focused mainly 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.

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

☆ We thank the editor, two anonymous referees, Kay Blaufus, Jochen Hundsdoerfer, Lisa Bruttel, Peter Mohr, Simeon Schudy and participants of the 2018 theem Conference, the 2017 arqus Conference, and the 2017 GfeW Conference for helpful comments and suggestions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author at: School of Business & Economics, Free University of Berlin, Thielallee 73, D-14195 Berlin, Germany.
E-mail addresses: martin.fochmann@fu-berlin.de (M. Fochmann), nfochmann@steuern.uni-hannover.de (N. Fochmann), martin.kocher@univie.ac.at (M.G. Kocher), nadja.mueller@wiso.uni-koeln.de (N. Müller).

https://doi.org/10.1016/j.jebo.2021.02.018

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), Korbel (2017), Mühlheußer 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 the group interaction induces a spillover 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 matters in the (non-)compliance decision as well.

The remainder of this paper is structured as follows: in Section 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 describes the details of our experimental design. We empirically analyze compliance behavior and treatment differences in Section 4. In Section 5, we study different types of decision makers and analyze the influence of individual preferences on group compliance. Arguments communicated in the group chats are examined in Section 6. Section 7 concludes the paper and draws implications for situations outside the laboratory.

2. Related literature

The general literature on differences between individuals and small groups as decision makers is large. 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).

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 of risk (Baker et al., 2008; Shupp and Williams, 2008; He et al., 2012).

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; 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),[1] 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.

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 laboratory 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; Lohse and Simon, 2018). However, not all studies find differences in unethical behavior between groups and individuals (Sutter, 2009; Azar et al., 2013; Mühlheußer 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 (Cohen et al., 2009). 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

[1] In contrast, Whyte (1993) and Rau (2015) observe stronger distortions in teams.

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.[2] 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*. In addition to a preference for being honest, Abeler et al. (2019) provide empirical evidence that a main motivation for truth-telling is a preference for being seen as honest. 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. Experimental design and predictions

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 1000 Lab-points.[3] The decision maker decides on how much of the actual income should be reported, and all integer values from 0 to 1000 are allowed. The company is requested to pay a corporate tax of 25% on the reported income.[4] With a probability of 30%, the report is audited. If the audit reveals that the reported income is less than the actual income of 1000, 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 = 1000 – 0.25 * reported income;

…if an audit occurs:

Company's after tax profit = 1000 – 0.25 * reported income – 2 * 0.25 * (1000 – reported income).

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 * company's after tax profit

We cannot exclude that the company framing has no effect on compliance behavior. We are not aware of any papers addressing such framing effects in the compliance literature, but, given results from related areas of research, the frame might lead to a general shift (increase or decrease) in tax compliance behavior. To ensure that such a level effect does not distort our analysis, we of course apply the same company framing in all our treatments, and our discussion focuses on the compliance differences between treatments and not on the interpretation of absolute compliance levels.

Assuming a risk-neutral decision maker, the payoff-maximizing strategy in expected value terms is to report zero income. In case of full compliance (i.e., reported income is 1000), the decision maker's payoff is 170 (= 20 + 0.2 * (1000 – 0.25 * 1000)), independent of whether the audit occurs. In case of full non-compliance (i.e., reported income is zero), the expected payoff is 190 (= 0.7 * (20 + 0.2 * 1000) + 0.3 * (20 + 0.2 * (1000 – 2 * 0.25 * 1000)). Any reported income between zero and 1000 yields an (expected) payoff between the two values. Obviously, the decision maker earns most if she or he does not comply and is not audited (220 = 20 + 0.2 * (1000)).

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 by sending text messages in an anonymous, five-minute chat on the computer screen, without being allowed to

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

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

[4] 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).

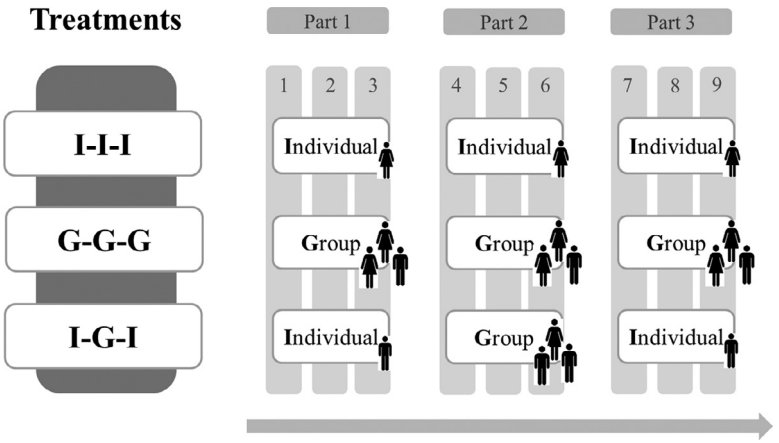


Fig. 1. Experimental design.
Note: In treatment I–I–I, the individual (I) setting is applied in all three parts. In treatment G–G–G, the group (G) setting is applied in all three parts. In treatment I–G–I, the individual setting is applied in the first part, the group setting in the second, and the individual setting again in the third part. Each part consists of three income reporting decisions.

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.[5]

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). Fig. 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 treatment 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.

3.3. Predictions

The innovation of our study is the combination of the honesty dimension and the risk dimension stemming from the auditing chance, in a context in which individuals or groups have to decide whether to behave honestly or not. Based on our literature review in Section 2, we discuss in the following the behavioral determinants that come into play in our experiment and derive our hypothesis. We focus on the two main aspects of our design.

Lying and cheating behavior. The literature on lying and cheating provides rather conclusive evidence for a dishonesty shift in groups. We expect to replicate this finding. The first two mechanisms mentioned in the literature review (*groups tend to be more strategical than individuals* and *restricted individual accountability of groups*) obviously also hold in our experiment. In particular, as the payoff-maximizing strategy for risk-neutral decision makers is to evade all taxes (see Section 3.1) and as group members might feel less responsible for their actions when they decide in a group, compliance should be lower in our group than in our individual settings. As we allow group members to chat with each other before their make their decisions, the third mechanism – *communication and norm perception* – matters in our study, too. Following the conclusive evidence from the literature we expect to find lower compliance in the group setting according to this mechanism as well. As group members' unethical behavior might not only increase their own payoff (at least in expected value terms), but also the payoff of other group members, the fourth mechanism – *other people benefit from unethical behavior* – should also lead to lower compliance in the group settings. An argument against a stronger inclination to behave unethically in groups are mechanisms related to *image concerns*. However, we argue that once the group agreed on evading taxes jointly during chat communication, image concerns do play an important role anymore. Hence, disregarding the risk aspect, we should observe lower compliance levels by groups than by individuals.

Risk-taking behavior. From the literature on tax evasion, we know that risk attitude and risk perception have a substantial influence on (risky) tax compliance decisions. There is conclusive evidence that a higher willingness to take risks is associ-

[5] There are different ways how to aggregate individual opinions to group decisions. Existing studies have implemented median rules, unanimity requirements (with different defaults), or voting procedures. Given our decision variable, the median rule seems most appropriate. There is little systematic evidence from this literature on potential effects of different decision rules; clearly, the choice is not completely innocuous. The comparison with the results in Kocher et al. (2018), who use a unanimity requirement, shows, however, that the decision rules might play a smaller role for the comparison of individual and group decisions in our specific context than one would perhaps expect.

ated with lower compliance (e.g., Dulleck et al., 2016; Fochmann and Wolf, 2019). Consequently, if the willingness to take risks was higher in groups, compliance should be lower in our group than in our individual settings. Yet, given the mixed results in the economics and psychology literature on a risky/cautious shifts in groups, a design-based discussion of arguments is necessary. Following the first mechanism – *groups take more rational decisions* – we expect to see a tendency for lower compliance in our group setting, as the rational – payoff-maximizing – strategy is to report zero income. The second mechanism discussed in the literature review (*social responsibility*) is a bit more difficult to assess. Given the chat in the group, we expect quite a lot of congruent proposals. If no agreement is reached, probably social responsibility is less of a determinant of behavior, as the fact that the group could not find consensus potentially reduces feelings of responsibility. Due to the expectation of many congruent proposals, we do not expect this mechanism to matter much in our context. Following the third mechanism (*conformism)*, risk-taking and consequently compliance can increase or decrease dependent on the initial tendencies and potential informal hierarchies of the group members. Thus, a directed prediction is not possible, without knowing the attitudes of participants. Yet, the design of our treatment I–G–I allows us to study this effect by analyzing effects from the composition of groups (see Section 5.2).

Main hypothesis. As almost all discussed arguments and mechanisms that are relevant for our experiment suggest that groups should be less compliant than individuals, we formulate the following directed hypothesis.

Hypothesis: Compliance is lower in the group (G) settings than in the individual (I) settings.
Consequently, compliance is lower…

a) *in each part of treatment G–G–G than in treatment I–I–I;*
b) *in the second part of treatment I–G–I than in treatment I–I–I;*
c) *in the first part and the third part of treatment G–G–G than in treatment I–G–I; and*
d) *in the second part compared to the first part and the third part within treatment I–G–I.*

Ex ante it looks as if the driver of the difference should be more the group interaction per se and less the fact that risk is perceived or taken differently in groups, given the existing results from related setups that introduce the two aspects separately. However, it is difficult to compare the size of effects across different settings. Hence, we do not want to make a strong claim here, and do not formulate this expectation as a formal hypothesis. Moreover, it is unclear whether the two aspects – unethical behavior and decision making 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 reinforce the dishonesty shift when a risk of detection is present. Again, we view this as an exploratory aspect of our study, and do not want to formulate it as a formal hypothesis.

3.4. Experimental protocol

At the beginning of each part, participants receive written instructions in which all part-related information are presented. The instructions are available in Appendix A1. In the instructions for the first part, subjects are informed that the entire experiment consists of three parts in total and that each part consists of three decisions. Furthermore, participants are informed that, at the end of the experiment, for each subject one out of the nine decision situations 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 that is reproduced in Appendix A2. 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 could have added to the moral component in decision making (see also, Section 3.1).

In the group setting, three subjects are randomly assigned to one group. This assignment was fixed for the relevant 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 own calculations. The latter allows subjects to automatically calculate, for example, the company's after tax profit and individual payoffs 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).

Table 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 attitude	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 morality	0 to 9; low tax morality = 0; high tax morality = 9	6.87
Positive reciprocity	0 to 10; low positive reciprocity = 0; high positive reciprocity = 10	8.22
Negative reciprocity	0 to 10; low negative reciprocity = 0; high negative reciprocity = 10	5.46
Fairness	0 to 10; low perceived fairness of tax and control system in experiment = 0; high perceived fairness of tax and control system in experiment = 10	6.41
Decision complexity	0 to 10; low perceived decision complexity in experiment = 0; high perceived decision complexity in experiment = 10	1.72
Joy	0 to 10; felt no joy during experiment = 0; felt high joy during experiment = 10	6.17
Anger	0 to 10; felt no anger during experiment = 0; felt high anger during experiment = 10	3.81
Fear	0 to 10; felt no fear during experiment = 0; felt high fear during experiment = 10	1.86
Guilt	0 to 10; felt no guilt during experiment = 0; felt high guilt during experiment = 10	1.57
Income	In Euro (monthly income after fixed costs)	324.10
Religious	Praying at least once a week = 1; elsewise = 0	22.22%
What if calculations	Number of what if calculations used before submitting reported income	0.85

Note: This table presents all variables of our experiment. The respective ex-post questionnaire is presented in Appendix A2.

3.5. 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 min (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 1 provides an overview of all our variables and presents descriptive results.

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 1000), we can use the absolute values of declared income as our main 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).

Fig. 2 shows averages of declared income for the three treatments (standard deviations in parentheses). 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 ($p = 0.552$).

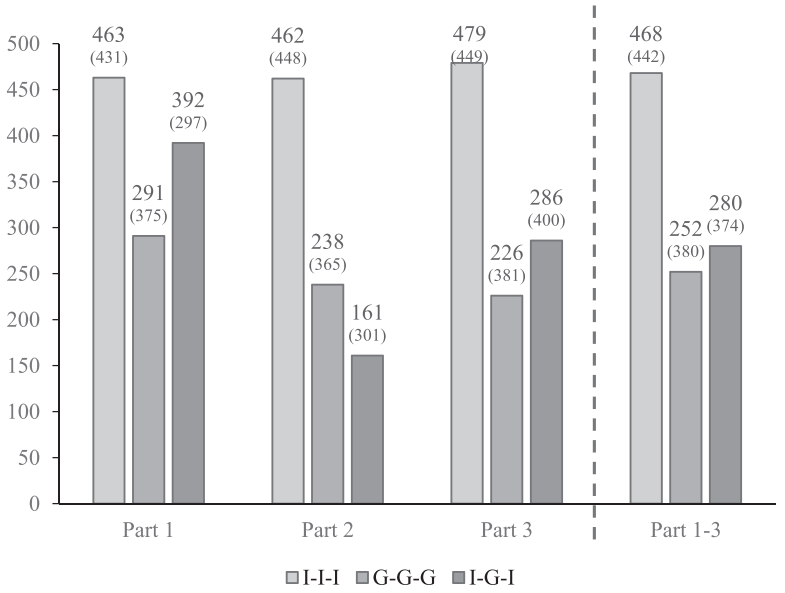


Fig. 2. Reported income.
Note: This figure shows mean reported income for each part and treatment (standard deviations in parentheses).

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.[6] The level is significantly different from part 3 in I–I–I (Mann–Whitney *U test*; 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 spillover effect on subsequent individual compliance: individual compliance is significantly lower after the group interaction.*

In Appendix A3.1, Fig. A3.1 shows the histograms for treatments I–I–I and G–G–G (pooled over all parts), and Fig. A3.2 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 levels in groups are extremely high, despite the fact that they are 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 (for more details see Appendix A3.2).

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.[7] 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 level and at the individual level.[8]

Table 2 reports the outcome for the comparison of treatments 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 significantly 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, and 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

[6] 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; $N = 69$).

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

[8] As robustness tests, we re-run all regressions as random-effects panel regressions (panel variable: subject ID, time variable: period) with standard errors clustered at the group level. All results are robust to this variation.

Table 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 subject			154.66***	153.82***
			(42.40)	(42.42)
Other individual controls	NO	NO	YES	YES
Constant	468.15***	463.01***	319.62**	284.46*
	(50.72)	(52.95)	(152.17)	(154.55)
No. of observations	1080	1080	960	960
No. of subjects	120	120	120	120
No. of independent groups	72	72	72	72
Wald tests:				
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). Statistical significance for the two dummies for the parts and the two interaction terms was checked with Wald tests, and the resulting *p*-values are reported in the bottom of the table.

*** $p \leq 0.01$.

** $p \leq 0.05$.

* $p \leq 0.1$.

treatments I–I–I and G–G–G in parts 2 and 3, respectively. 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 Fig. 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 models 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 1. We show the coefficient of the dummy variable "female subject" in Table 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.[9] Again, we observe 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.[10]

In Table 3, we use the same approach and specifications to analyze the differences between treatments I–I–I and I–G–I. In models 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

[9] The complete set of all regression results are presented in Appendix A4.1.

[10] 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
Treatment I–I–I vs. I–G–I: multi-level mixed effects linear regressions (dependent variable: reported 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 subject			178.48***	177.84***
			(49.32)	(49.07)
Other individual controls	No	No	Yes	Yes
Constant	468.15***	463.01***	342.57	361.21
	(47.39)	(50.48)	(222.89)	(224.33)
No. of observations	1053	1053	936	936
No. of subjects	117	117	117	117
No. of independent groups	71	71	71	71
Wald tests:				
Part 2 = Part 3		$p = 0.5806$		$p = 0.5249$
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. Statistical significance for the two dummies for the parts and the two interaction terms was checked with Wald tests, and the resulting p-values are reported in the bottom of the table. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

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 that 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 spillover effect in treatment I–G–I (Result 2). In line with the regression results presented in Table 3, we find a positive and significant effect of "female subject" and a negative and significant effect of "last period audit".

As a robustness check, we re-run our regressions with a binary dependent variable "honest report" (1 if reported income is 1000, and 0 otherwise). Although this variable does not take the magnitude of a misreport into account, it is worth to be studied, as it measures compliance at the extensive margin. Regression results are displayed in Appendix A4.2 (Tables A4.3 and A4.4). Most importantly, results are robust to this modeling variation, and compliance levels are significantly lower in the group than in the individual setting again. Thus, Result 1 is also observed with the binary dependent variable. However, Result 2 is not confirmed. When comparing treatments I–I–I and I–G–I, we do not see a significant interaction term "Part 3 X Treatment I–G–I" (see models 2 and 4 in Table A4.4); compliance levels in part 3 do not differ across the two treatments. Thus, in case of the binary dependent variable, there is no evidence that group interaction induces a spillover effect on subsequent individual compliance. From the evidence we can take two conclusions for treatment I–G–I. First, individuals who are fully compliant in part 1 are also fully compliant in part 3. Second, individuals who are not fully compliant in part 1 become less compliant in part 3 after the group interaction. Thus, the group interaction has a spillover effect only on individuals who are initially not fully compliant, but not on individuals who are initially fully compliant.

5. Treatment I–G–I: types of decision makers and group composition

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 in the following whether we find pat-

Table 4
Types of decision makers in treatment I–G–I.

Types	All subjects (69 subjects)				Men only (35 subjects)				Women only (34 subjects)			
	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%			0	0%			1	1%
	1	1%			0	0%			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.

terns 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.[11] Examples are shown in Appendix A5.

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.[12] The entire distribution of subjects in our categorization of different decision maker types is displayed in Table 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.

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.

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 task (Holt and Laury, 2002), and the number of female group members. Table 5 presents the regression results. Remarkably, the coefficient of mean reported income from part 1 is positive and significant. This suggests that groups tend to be more compliant if the group consists of more compliant individuals (i.e., who revealed higher compliance in part 1). This provides some evidence for a conformism effect. Furthermore, the number of female

[11] As robustness checks, we also used transition levels of 150 and 200 Lab-points. All results are qualitatively robust to this variation. Indeed, the definition of the transition level is crucial for the categorization of individuals and the subsequent analysis. One the one hand, if the transition level is set too low (e.g., 20) even small changes are interpreted as a transition, although no substantial change in behavior might have occurred. On the other hand, if the transition level is set too high (e.g., 500), relevant changes in compliance behavior might not be recognized and consequently would not be interpreted as such. As we observed spikes for round values in the histograms, we decided to take 100, 150 and 200 as potential transition levels. We think that this is a meaningful compromise for balancing the risk in both directions.

[12] In a public goods experiment with taxes, 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 run and the long run).

Table 5
Group composition in treatment I–G–I (dependent variable: mean reported income in part 2).

	Individual compliance in part 1 Model 1	Individual risk preferences Model 2	Gender 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	237.82	−8.91
	(96.52)	(222.62)	(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). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

group members shows up significantly in the regressions.[13] The coefficient for risk has the correct sign, but it is far from being significant.

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.[14] 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 relevant message independently. For each variable of interest, the median value of all three coders determined the coding that was finally used in our analysis. In total, we have 47 groups engaged in 285 chats.[15]

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%).[16] 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 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."). Fig. 3 displays the share of chats, in which these arguments are brought forward. The most frequent type of argument used is associated with risk, for both encouraging and discouraging compliance.[17]

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.

[13] As already mentioned, women being more compliant than men is a robust finding in the tax compliance literature. At the individual level, we found support for this result (see our regressions in tables 2 and 3). Here, we provide additional evidence that group composition matters, in addition to the difference in compliance on the individual level.

[14] The complete codebook with all variables of interest can be found in appendix A6.

[15] 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 à three decision situations; 9 x 24 groups = 216 chats). Each chat lasts five minutes.

[16] In addition, we analyzed the number of messages sent with arguments used to encourage compliance or non-compliance separately for the frequent types of decision makers in treatment I–G–I according to the classification from section 5. We found that arguments for non-compliance were brought forward by all types of decision makers and to a very similar extent. Remarkably, almost half of all compliance messages (48%) were brought forward by the decrease-increase-type, whereas the decrease-constant-type sent nearly no such messages. The latter result should be interpreted with caution, as the absolute number of sent compliance messages per type is very small.

[17] Result 4 also holds when we analyze the chat separately for all types of decision makers in treatment I–G–I.

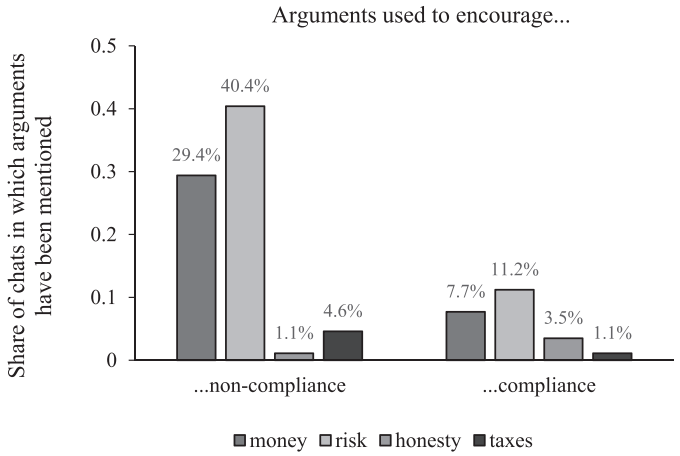


Fig. 3. Arguments used in group chats.

Table 6
Group chats: linear regressions with random effects (dependent variable: income reported by group).

Dummy variables (arguments used)	Treatment I–G–I and G–G–G			
	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 with income reported by the group finally in a period as dependent variable (regression coefficients, standard errors in parentheses). Since groups face repeated decisions, we run linear regression models with random effects, where the period is the time variable and the group's identity number is the cross-sectional variable. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

Table 6 displays linear regression results with income finally declared by the group in a period as dependent variable. As independent variables, we use dummy variables indicating whether a specific type of argument is mentioned in a chat.[18] Whereas models 1 and 2 consider the general use of arguments in favor of compliance or non-compliance, models 3 and 4

[18] 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 variables.

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 the one for non-compliance. The regression coefficients differ significantly from each other (Wald-test, $p<0.001$, two-tailed). Thus, arguments for compliance have a 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 related to risk.

In Appendix A7, we analyze how communication changes over time. Our main finding is that the exchange of arguments is reduced over time, but does not stop. In many cases, the group discussion in later periods is about sticking to the previous strategy.

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 in 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 choose less honestly. Importantly, the risk dimension is the most relevant aspect in the decision making procedure within groups. It is most often discussed in the group chats, and it has an effect on the outcome of the declared group income.

We also find some evidence on a spillover 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. Remarkably, the spillover effect is only observed for individuals who are initially not fully compliant, but not for individuals who are fully compliant in the first individual decision. Obviously, depending on the decision rule, different compositions of groups with regard to the individual inclination to comply can lead to very different group outcomes and perhaps even different spillover effects.

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. 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 could look at different detection probabilities to further corroborate these first results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.02.018.

Appendix

A1. Instructions

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

A1.1. General instructions

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

The experimental study consists of 2 experiments in which you have the opportunity to earn money. Before each experiment, you will receive instructions describing each experiment. Then the experiment starts. After completing the second experiment, you will receive a payout (in addition to the participation fee) which depends on the results of both experiments.

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

Before we start, we would like to draw your attention to a few important points.

- Please note that you are neither allowed to communicate with other participants nor allowed to leave your desk during both experiments. Please do not look at what other participants are doing.
- Please turn off your mobile phone and store it in your bag.
- Please read the instructions thoroughly.
- It is important that you understand the instructions. Therefore, please do not be afraid to ask questions. If you have any questions, please raise your hand. We will then come to you to answer your questions. Please do not ask questions aloud.
- You can write and make markings on the instructions.
- The calculator and the pen that are lying in front of you can be used.
- Please do not take the instructions home, but return them to us at the end of the study.

Before the first experiment starts, we ask you to fill in a short questionnaire on your computer.

After that the instructions for the first experiment will be displayed on your monitor.

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

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

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

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

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

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

A1.3. Instructions for main experiment

A1.3.1. Instructions for treatment I–I–I

A1.3.1.1. Instructions for Part 1.

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Corporate employee and corporate income

Imagine you are the employee of a company. Your task is to file the tax return for the company.

As an employee, you receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25%. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income you want to declare (in the amount of 1000 lab-points). All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25% of the declared corporate income:

tax = 0,25 × *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:

$$\text{unpaid tax} = 0,25\text{x}\left(\underbrace{1000}_{\text{actual corporate income}} - \text{declare corporate income}\right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20% of the company's success. Your personal payout will be as follows:

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

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

When deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting company's success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions at your computer. Answering these questions is only a check of your understanding and is not payout relevant.

A1.3.1.2. Instructions for Part 2.

The second part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. The second part of the experiment again consists of 3 periods in which you make one decision each.

Corporate employees and corporate income

No changes to the first part of the experiment.

Continue to imagine you are an employee of a company. Your task is to file the tax return for the company.

As in the first part of the experiment, you as an employee receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

No changes to the first part of the experiment.

Therefore, in each period, a tax is again imposed at a rate of 25%.

The amount of tax to be paid by the company continues to be based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income (which is 1000 lab-points) you want to declare. All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25% of the declared corporate income:

tax = 0,25 × *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:

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

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

No changes to the first part of the experiment.

It therefore continues to apply that when deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

A1.3.1.3. Instructions for Part 3.

The third part of the experiment is identical to the first and second part of the experiment. This means that you make the same decisions as in the first and second part. The third part of the experiment again consists of 3 periods in which you make one decision each.

A1.3.2. Instructions for treatment I–G–I

A1.3.2.1. Instructions for Part 1

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Corporate employee and corporate income

Imagine you are the employee of a company. Your task is to file the tax return for the company.

As an employee, you receive a fixed remuneration of 20 lab-points. In addition, you receive a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25%. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income declared by you in the tax return of the company. To do this, you simply determine how much of the actual corporate income you want to declare (in the amount of 1000 lab-points). All integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The tax payable amounts to 25% of the declared corporate income:

tax $=$ 0,25 $\times$ *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:

$$\text{unpaid tax} = 0,25 \text{ times} \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20% of the company's success. Your personal payout will be as follows:

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

Please note: Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.

After each period, you will receive information about whether an audit has been carried out or not. In addition, you will receive an overview of all important data as well as your personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

When deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting company's success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions at your computer. Answering these questions is only a check of your understanding and is not payout relevant.

A1.3.2.2. Instructions for Part 2

The second part of the experiment is identical to the first part of the experiment. The only exception is that you now make your decisions in a triad. Your remuneration therefore also depends on the decisions of other participants.

The second part of the experiment again consists of 3 periods in which you make one decision each.

Group

Together with 2 other, randomly selected participants, you form a triad that stays together during the second part of the experiment. Each of these 3 group members makes the same decisions.

Corporate employees and corporate income

Imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

As in the first part of the experiment, each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company's success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

Therefore, in each period, a tax is again imposed at a rate of 25%.

The amount of tax to be paid by the company is based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (which is 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by size (from small to large). This also means that if two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25% of the declared corporate income:

tax = 0,25 × *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 min in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

No changes to the first part of the experiment.

It therefore continues to apply that the provided information on the corporate income is audited with a probability of 30%. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, as in the first part of the experiment, the company must pay a fine equal to the unpaid tax.

Therefore, the company's success continues to result in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

No changes to the first part of the experiment.

Your personal payout in a period continues to consist of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20% of the company's success. Your personal payout will be as follows:

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

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

While chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

Information about the chat

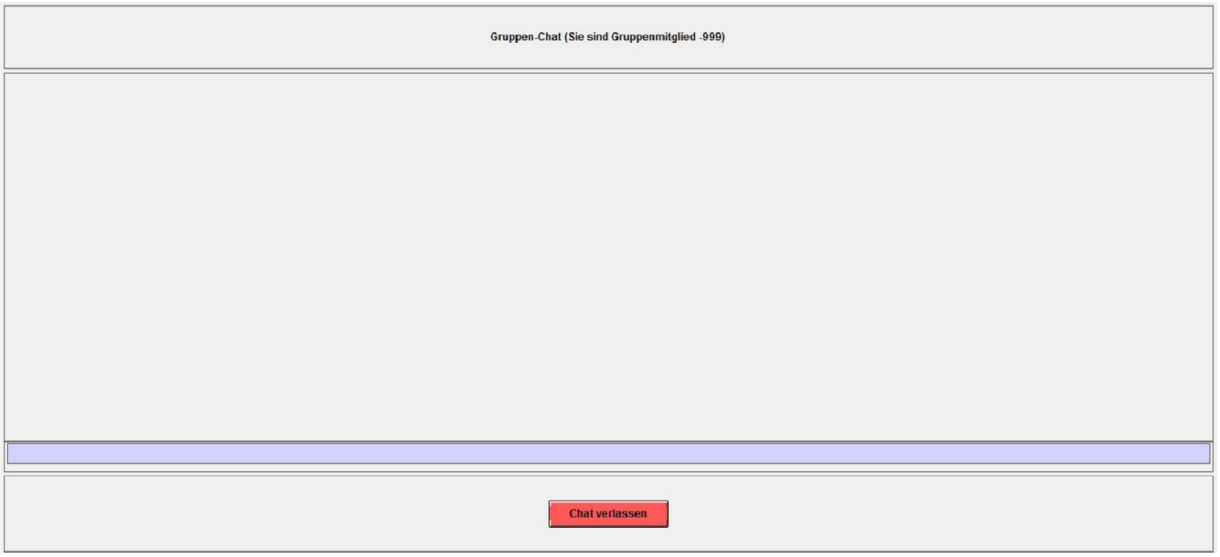
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

You have 5 min to exchange information. The group discussion ends after 5 min or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do not want to leave the chat, you can click on the button "Back". After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the "enter" key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the "enter" key, that is, when your message appears in the gray box.

A1.3.2.3. Instructions for Part 3

The third part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. Please note, therefore, that you make the decisions *on your own* and not in a group anymore. The third part of the experiment again consists of 3 periods in which you make one decision each.

A1.3.3. Instructions for treatment G–G–G

A1.3.3.1. Instructions for Part 1

General information

The second experiment consists of 3 parts. The decision situations in the 3 parts are basically identical. Before each part of the experiment, you will receive instructions explaining that part of the experiment. Each part of the experiment consists of 3 periods in which you make one decision each. In total, you make 9 decisions. At the end of the second experiment, one of the 9 decisions will be randomly selected and paid out. How much money you earn depends on your decisions, the decisions of other participants, and on chance. These instructions explain to you how to earn money in this experiment. Therefore, read the following paragraphs thoroughly.

For simplification purposes, this experiment does not calculate in euro amounts, but in lab-points. One lab-point is exactly 10 euro cents. That means 100 lab-points are exactly 10 euros.

Group

Together with 2 other, randomly selected participants, you form a triad that stays together during the first part of the experiment. Each of these 3 group members makes the same decisions.

Corporate employees and corporate income

Imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

Each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company's success. How exactly your personal payout will be calculated is explained below.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

In each period, a tax is imposed at a rate of 25%. The tax revenues will be used to fund future experiments.

The amount of tax to be paid by the company is based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (in the amount of 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by

size (from small to large). This also means that if two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25% of the declared corporate income: tax = 0,25 × *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 min in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

With a probability of 30%, the provided information on the corporate income is audited. With the counter-probability of 70%, the information is not audited. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, the company must pay a fine equal to the unpaid tax.

tax repayment = unpaid tax

Fine = unpaid tax

The unpaid tax is:

$$\text{unpaid tax} = 0,25 \times \left(\underbrace{1000}_{\text{actual corporate income}} - \text{declared corporate income} \right)$$

The company's success results in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

Your personal payout in a period consists of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20% of the company's success. Your personal payout will be as follows:

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

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

Since the second experiment consists of 3 parts, each of which consists of 3 periods, you make decisions in 9 periods. After making decisions in all 9 periods, *one* period is randomly selected by the computer at the end of the second experiment. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

While chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen). Among other things, this will show you the resulting corporate success as well as your personal payout, both in the event that no audit is carried out and that an audit is carried out. In addition, you can use the calculator at your workplace for your own calculations.

Before the second experiment starts, you are asked to answer some questions on your computer. Answering these questions is only a check of your understanding and is not payout relevant.

Information about the chat

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

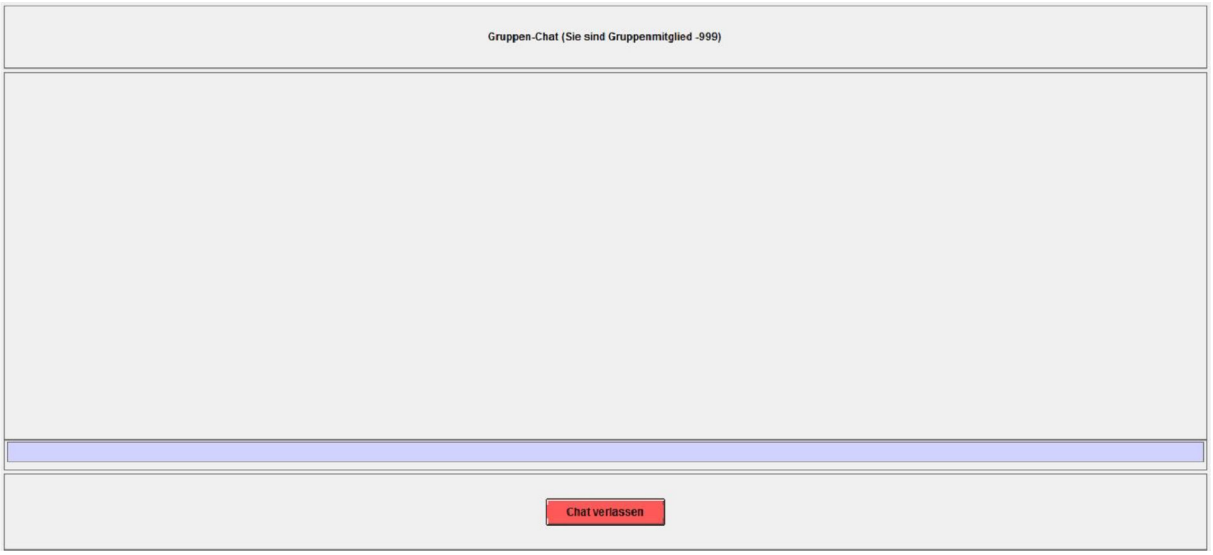
You have 5 min to exchange information. The group discussion ends after 5 min or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do

not want to leave the chat, you can click on the button “Back”. After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the “enter” key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the “enter” key, that is, when your message appears in the gray box.

A1.3.3.2. Instructions for Part 2

The second part of the experiment is identical to the first part of the experiment. This means that you make the same decisions as in the first part. The second part of the experiment again consists of 3 periods in which you make one decision each.

Group

Please note that your triad consists of the same group members as in the first part of the experiment and that you therefore interact again in the second part of the experiment with the same participants. Each of the 3 group members makes the same decisions again.

Corporate employees and corporate income

No change to the first part of the experiment.

Continue to imagine you and the other two members of your group are employees of a company. Your common task is to file the tax return for the company.

As in the first part of the experiment, each employee receives a fixed remuneration of 20 lab-points. In addition, each employee receives a variable remuneration, which depends on the company’s success.

In each period, the company has earned a corporate income of 1000 lab-points.

Tax return of the company

No changes to the first part of the experiment.

Therefore, in each period, a tax is again imposed at a rate of 25%.

The amount of tax to be paid by the company continues to be based on the corporate income, which your group declares in the tax return of the company. The group decides by vote on the amount of the declared corporate income. For this purpose, each individual group member makes a personal proposal of how much of the actual corporate income (which is 1000 lab-points) should be declared. As a proposal all integer values between 0 and 1000 are possible, whereby the numbers 0 and 1000 can also be chosen. Please note: The declared corporate income can therefore be equal to or less than the actual corporate income, but not higher.

The median of the proposals of the three group members determines the amount of the declared corporate income of your group in this period. The median is the value that stands at the middle (central) location when sorting the values by

size (from small to large). This also means that if two or three group members propose the same value, this proposed value is the median. Please note that the median is not the same as the mean. After each member of the group has made his binding proposal, the median is automatically determined and the amount of the declared corporate income specified.

The tax payable is 25% of the declared corporate income:

tax = 0,25 × *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 min in a chat. More information about the chat can be found on the last page of these instructions.

Audit of tax declaration and corporate success

No changes to the first part of the experiment.

It therefore continues to apply that the provided information on the corporate income is audited with a probability of 30%. If there is an audit and the declared corporate income does not coincide with the actual corporate income, the company has to repay the unpaid tax. In addition, as in the first part of the experiment, the company must pay a fine equal to the unpaid tax.

Therefore, the company's success continues to result in the case of an audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax} - \text{tax repayment} - \text{fine}$$

The company's success results in the case of no audit as follows:

$$\text{company's success} = \underbrace{1000}_{\text{actual corporate income}} - \text{tax}$$

Your personal payout in a period

No changes to the first part of the experiment.

Your personal payout in a period continues to consist of two components. On the one hand, you receive a fixed remuneration of 20 lab-points. On the other hand, you receive a variable remuneration which depends on the company's success. The variable remuneration amounts to 20% of the company's success. Your personal payout will be as follows:

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

Please note:

- Since your personal payout depends on the company's success, it also depends on the tax and (possible) fine paid by the company.
- Each member of your group receives the same payout.

After each period, each group member will receive information about whether an audit has been carried out or not. In addition, each member receives an overview of all important data as well as the personal payout.

Payout

No changes to the first part of the experiment.

It therefore continues to apply that at the end of the second experiment (after making decisions in all 9 periods), *one* period is randomly selected by the computer. The payout of this period is converted into euros and will then be paid out to you in cash.

Final information

No changes to the first part of the experiment.

It therefore continues to apply that while chatting and deciding how much corporate income you want to declare, you have the option to run trial calculations on your computer (lower half of the screen).

Information about the chat

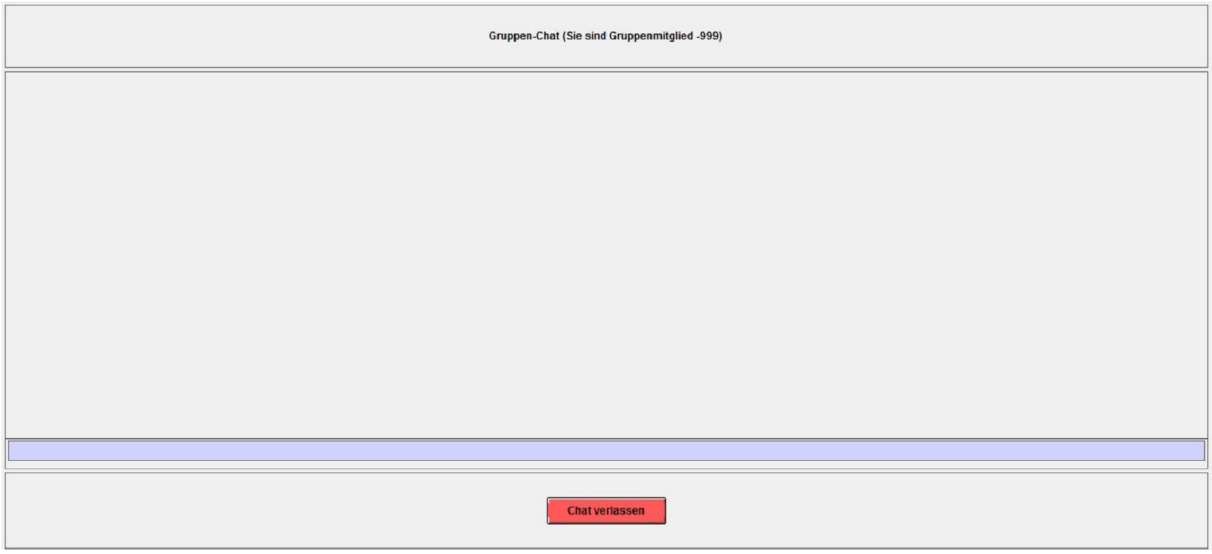
You have the option of communicating with the other two members of your group through a chat to discuss the proposal on the amount of declared corporate income that each group member will subsequently enter.

You have 5 min to exchange information. The group discussion ends after 5 min or as soon as all 3 group members have clicked the button "Leave Chat". If 1 or 2 group members click on the button, the chat will continue until either all group members have clicked on the button or the time has expired. If you have clicked on the button "Leave Chat", but do not want to leave the chat, you can click on the button "Back". After the group discussion, each member makes a binding proposal on the amount of the declared corporate income.

Basically, the content of the communication is open, but it is forbidden to share personal information. Personal data is: name, age, gender (please always use gender-neutral terms), subject (this includes the mentioning of specific lecturers, courses or course descriptions, that allow for identification of the subject) or similar topics that could identify you (e. g. your cabin number or row). Furthermore, it is prohibited to agree on side payments within your group. If you violate these rules, you will be excluded from the experiment and will not receive a payout for the entire experiment.

During the given time each group member can send as many messages as he likes. Each of your messages automatically appears on the screen of the other two group members. Messages to a single person are not possible.

The screen with the chat will look like this:



To send a message, click on the purple box, type in your message and press the “enter” key. After that, your message will appear in the gray box above. This procedure allows you to send as many messages as you want. The other group members see your messages only when you hit the “enter” key, that is, when your message appears in the gray box.

A1.3.3.3. Instructions for Part 3

The third part of the experiment is identical to the first and second part of the experiment. This means that you make the same decisions as in the first and second part. The third part of the experiment again consists of 3 periods in which you make one decision each.

Please note that your triad consists of the same group members as in the first and second part of the experiment and that you therefore interact again in the third part of the experiment with the same participants.

A2. 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

A3. Histograms and in-group coordination

A3.1. Histograms

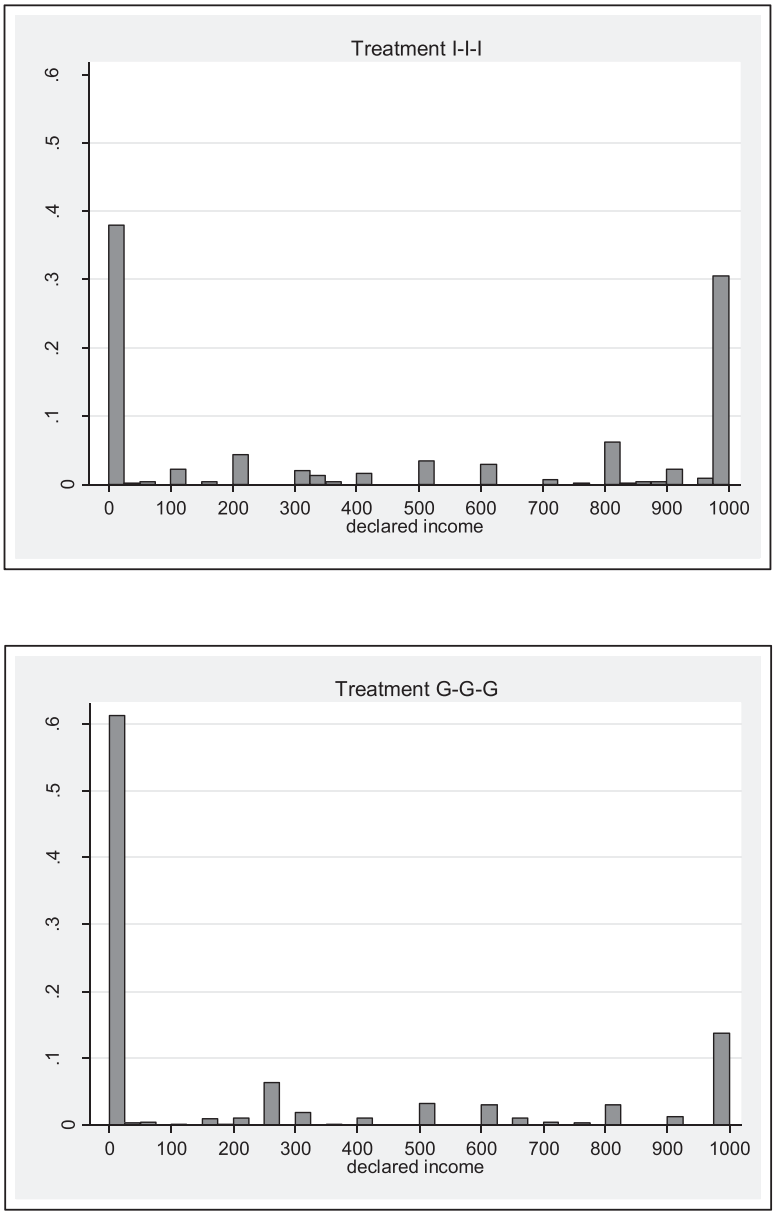


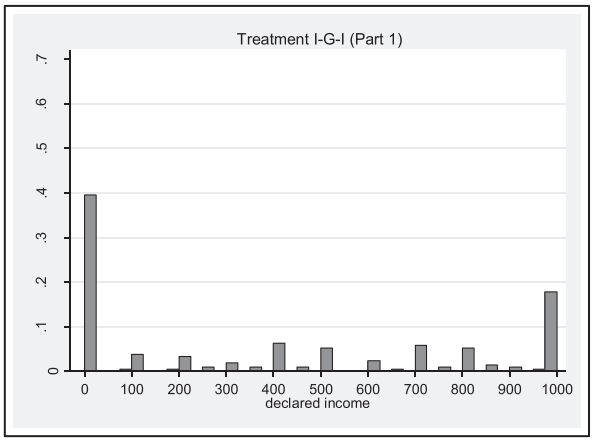
Fig. A3.1. Histograms for treatments I–I–I and G–G–G (data pooled over all parts).

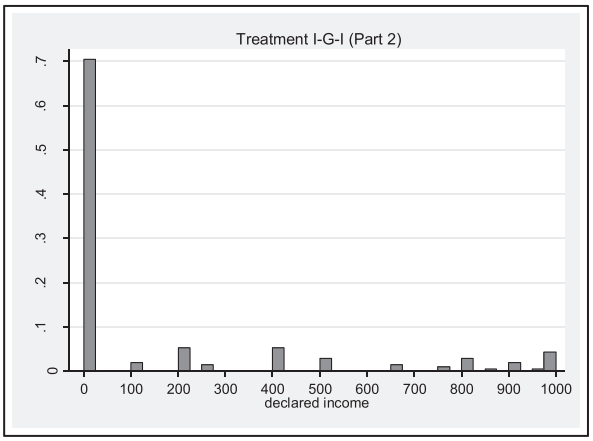
.

A3.2. In-group coordination

To measure in-group coordination, we divided the number of congruent votes by the total number of votes. A vote is recognized as a congruent vote if the declared income proposed by an individual equals the income finally declared by her/his group. We calculated rates for each period and treatment (see Table A3.1).

In all periods, we observe that coordination is very high (almost 90%, on average). However, in treatment G–G–G, coordination rates are a bit lower in the first two periods (period 1: 74%, period 2: 86%), before coordination continues on an almost constant rate of 92%. In treatment I–G–I (with the group setting in periods 4–6 only), coordination is almost constant in all periods (around 88%). One reason why coordination is higher in the first group periods in I–G–I than in G–G–G





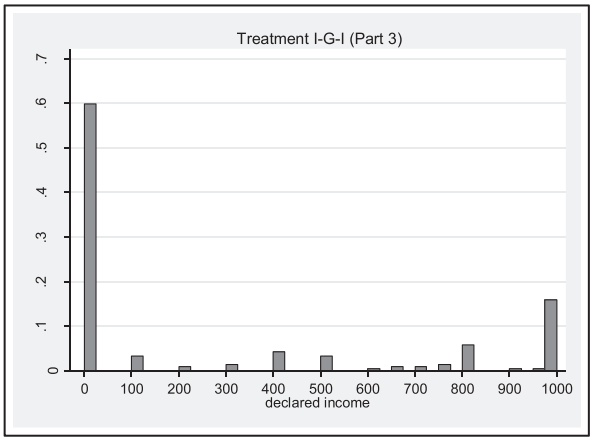


Fig. A3.2. Histograms for each part of treatment I–G–I.

Table A3.1
In-group coordination rates.

Period	Treatment I–G–I	Treatment G–G–G
1		74%
2		86%
3		92%
4	88%	92%
5	88%	96%
6	87%	92%
7		93%
8		92%
9		92%
1–9	88%	90%

might be that subjects had been able to familiarize themselves with the task in the three preceding periods of the individual setting in I–G–I (periods 1–3).

A4. Regressions

A4.1. Complete set of regression results (dependent variable: declared income)

This section presents the complete set of regression results. Table A4.1 corresponds to Tables 2 and A4.2 to Table 3.

A4.2. Regressions with binary dependent variable

Regressions presented in Tables 2 and 3 are re-run with the binary variable honest report as dependent variable. We use logistic panel regressions with random effects, where the subject's identity number is the cross-sectional variable. Table A4.3

Table A4.1
(Corresponds to Table 2): Treatment I–I–I vs. G–G–G: multi-level mixed effects linear regressions (dependent variable: declared income).

	Treatment I–I–I vs. G–G–G			
	Model 1	Model 2	Model 3	Model 4
Treatment G–G–G	−216.60**	−172.20**	−224.38***	−176.22**
	(84.81)	(87.06)	(83.39)	(87.73)
Part 2		−0.60		29.98
		(26.36)		(29.17)
Part 2 × Treatment G–G–G		−51.98		−46.38
		(34.03)		(36.48)
Part 3		16.03		49.13
		(26.36)		(30.10)
Part 3 × Treatment G–G–G		−81.22**		−71.61*
		(34.03)		(37.18)
Last period audit			−60.15***	−59.92***
			(20.31)	(20.17)
Period (1 to 3) within part			7.56	8.50
			(8.74)	(9.11)
Female subject			154.66***	153.82***
			(42.40)	(42.42)
Risk attitude			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)
Positive 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	1080	1080	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 × Treatment G–G–G = Part 3 × 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). Statistical significance for the two dummies for the parts and the two interaction terms was checked with Wald tests, and the resulting *p*-values are reported in the bottom of the table.

*** $p \leq 0.01$.
** $p \leq 0.05$.
* $p \leq 0.1$.

Table A4.2
(Corresponds to Table 3): Treatment I–I–I vs. I–G–I: multi-level mixed effects linear regressions (dependent variable: declared income).

	Treatment I–I–I vs. I–G–I			
	Model 5	Model 6	Model 7	Model 8
Treatment I–G–I	−188.49**	−71.16	−187.90**	−66.25
	(73.76)	(77.15)	(79.06)	(83.33)
Part 2		−0.60		35.20
		(30.10)		(32.16)
Part 2 × Treatment I–G–I		−229.76***		−217.66***
		(39.19)		(40.88)
Part 3		16.03		53.01
		(30.10)		(32.46)
Part 3 × Treatment I–G–I		−122.23***		−107.36***
		(39.19)		(40.82)
Last period audit			−42.16**	−57.21***
			(18.82)	(18.21)
Period (1 to 3)			22.17**	13.75
			(10.27)	(10.41)
Female subject			178.48***	177.84***
			(49.32)	(49.07)
Risk attitude			−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)
Positive 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 × Treatment I–G–I = Part 3 × 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). Statistical significance for the two dummies for the parts and the two interaction terms was checked with Wald tests, and the resulting *p*-values are reported in the bottom of the table.

*** $p \leq 0.01$.
** $p \leq 0.05$.
* $p \leq 0.1$.

Table A4.3
Treatment I–I–I vs. G–G–G: logistic panel regressions with random effects (dependent variable: honest report dummy).

	Treatment I–I–I vs. G–G–G			
	Model 1	Model 2	Model 3	Model 4
Treatment G–G–G	−3.18***	−3.05***	−3.71***	−3.58**
	(0.97)	(1.09)	(1.28)	(1.54)
Part 2		0.19		0.50
		(0.43)		(0.61)
Part 2 × Treatment G–G–G		−0.75		−0.75
		(0.76)		(0.96)
Part 3		0.46		0.75
		(0.43)		(0.65)
Part 3 × Treatment G–G–G		0.01		0.17
		(0.71)		(0.95)
Last period audit			−0.76**	−0.72*
			(0.39)	(0.39)
Period (1 to 3) within part			0.35	0.43*
			(0.22)	(0.23)
Female subject			−0.20	−0.24
			(1.34)	(1.41)
Other individual controls	No	No	Yes	Yes
Constant	−3.14***	−3.45***	4.48	4.09
	(0.80)	(0.86)	(5.68)	(6.07)
No. of observations	1080	1080	960	960
No. of subjects	120	120	120	120
No. of independent groups	72	72	72	72
Wald test:				
Part 2 = Part 3		$p = 0.5250$		$p = 0.6083$
Part 2 × Treatment G–G–G = Part 3 × Treatment G–G–G		$p = 0.3131$		$p = 0.2658$

Note: In this table, the results of logistic panel regressions with random effects are presented with honest report as dependent variable (regression coefficients, standard errors in parentheses).
*** $p \leq 0.01$.
** $p \leq 0.05$.
* $p \leq 0.1$.

Table A4.4
Treatment I–I–I vs. I–G–I: logistic panel regressions with random effects (dependent variable: honest report dummy).

	Treatment I–I–I vs. I–G–I			
	Model 1	Model 2	Model 3	Model 4
Treatment I–G–I	−2.24***	−1.45*	−2.71***	−1.87*
	(0.65)	(0.76)	(0.82)	(1.03)
Part 2		0.17		0.45
		(0.41)		(0.54)
Part 2 × Treatment I–G–I		−4.64***		−5.23***
		(0.94)		(1.10)
Part 3		0.41		0.65
		(0.41)		(0.56)
Part 3 × Treatment I–G–I		−0.64		−0.42
		(0.57)		(0.72)
Last period audit			−0.19	−0.68**
			(0.30)	(0.34)
Period (1 to 3)			0.39**	0.40**
			(0.17)	(0.21)
Female subject			0.87	1.15
			(0.86)	(0.97)
Other individual controls	No	No	Yes	Yes
Constant	−1.85***	−2.20***	0.83	0.77
	(0.48)	(0.58)	(3.65)	(4.14)
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		$p = 0.5478$		$p = 0.6512$
Part 2 × Treatment I–G–I = Part 3 × Treatment I–G–I		$p = 0.0000$		$p = 0.0000$

Note: In this table, the results of logistic panel regressions with random effects are presented with honest report as dependent variable (regression coefficients, standard errors in parentheses).
*** $p \leq 0.01$.
** $p \leq 0.05$.
* $p \leq 0.1$.

(corresponds to Table 2) compares treatments I–I–I and G–G–G and Table A4.4 (corresponds to Table 3) compares treatments I–I–I and I–G–I.

A5. Types of decision makers in treatment I–G–I

Figs. A5.1–A5.3 display examples for the most frequent types of decision makers observed in our experiment.

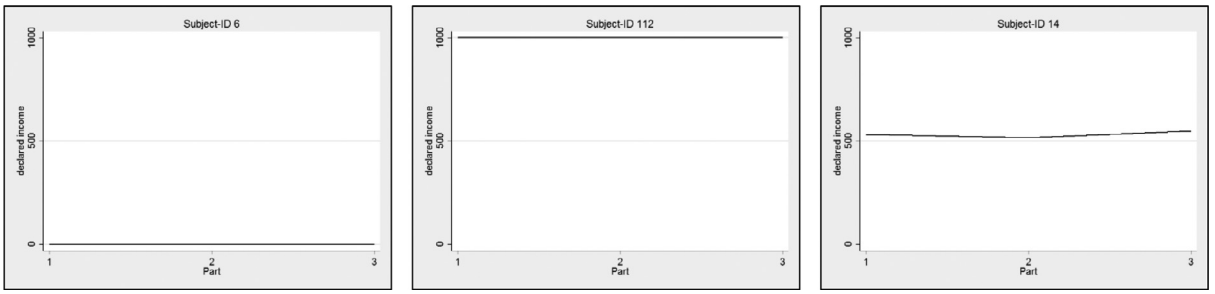


Fig. A5.1. Examples of constant-constant-types.

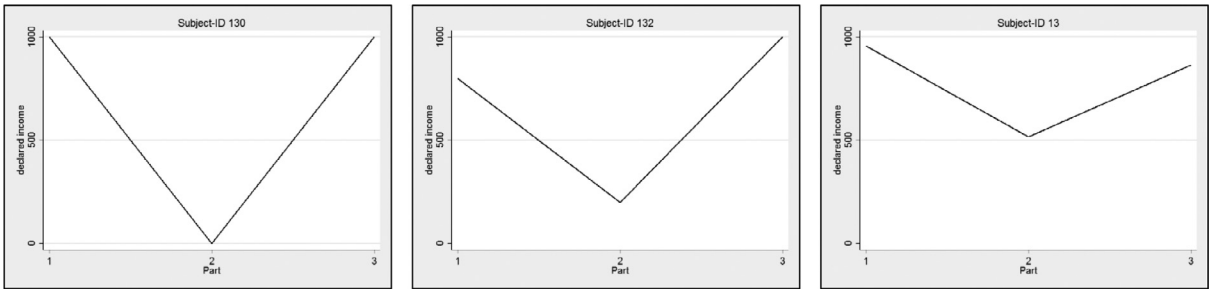


Fig. A5.2. Examples of decrease-increase-types (V-shape).

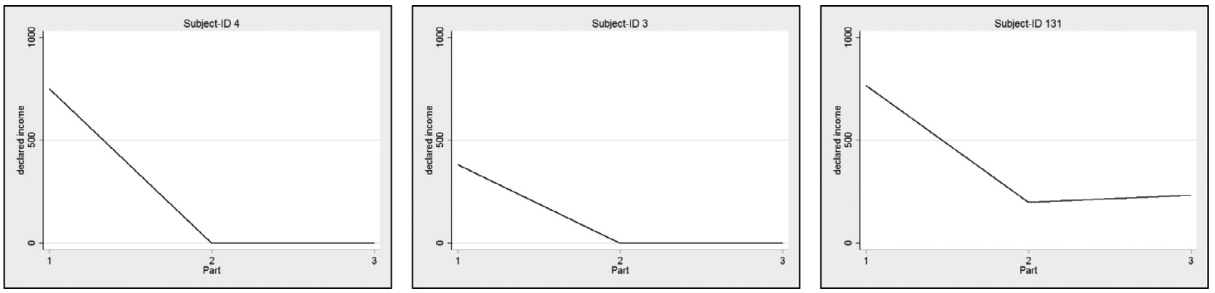


Fig. A5.3. Examples of decrease-constant-types (L-shape).

A6. 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"

A7. Communication over time

In this appendix, we analyze how communication changes over time. For this analysis, we additionally consider the argument categories 'keep strategy' (e.g., "Never change a running system.") and 'change strategy'.

In the first step, we look at the total number of categorized arguments. In treatment I–G–I, we categorized 68 arguments in the first (group setting) period, 65 in the second, and 49 in the third. A similar pattern is observed in treatment G–G–G: in the first part we categorized 85 arguments per period on average, 45 per period in the second part, and 32 in part 3. Consequently, we find that the number of exchanged arguments becomes smaller over time, but the discussion does not stop.

In the second step, we look at the arguments communicated. Figs. A7.1 and A7.2 show the share of chats, in which the arguments are brought forward, separated for the first three periods with group interaction (treatment I–G–I: periods 4–6; treatment G–G–G: periods 1–3). As discussed before, money and risk arguments are predominant. However, from the second period on, 'keep strategy' is the argument used most frequently to encourage non-compliance (Fig. A7.1). A similar result is observed for arguments encouraging compliance (Fig. A7.2), but in period three money and risk are again the arguments used most frequently. In both figures, we observe that the occurrence of 'change strategy' is negligible. The same pattern is observed when all three parts of treatment G–G–G are considered (see Figs. A7.3 –A7.5).

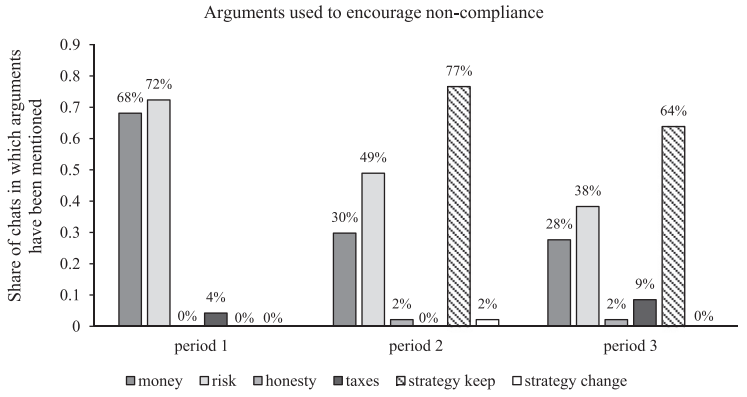


Fig. A7.1. Arguments used to encourage non-compliance in the first three periods with group interaction (treatment I–G–I: periods 4–6; treatment G–G–G: periods 1–3).

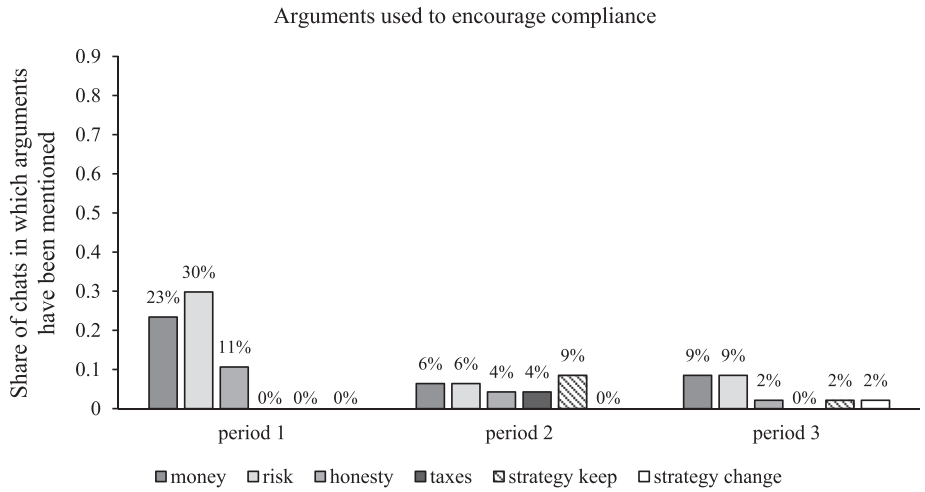


Fig. A7.2. Arguments used to encourage compliance in the first three periods with group interaction (treatment I–G–I: periods 4–6; treatment G–G–G: periods 1–3).

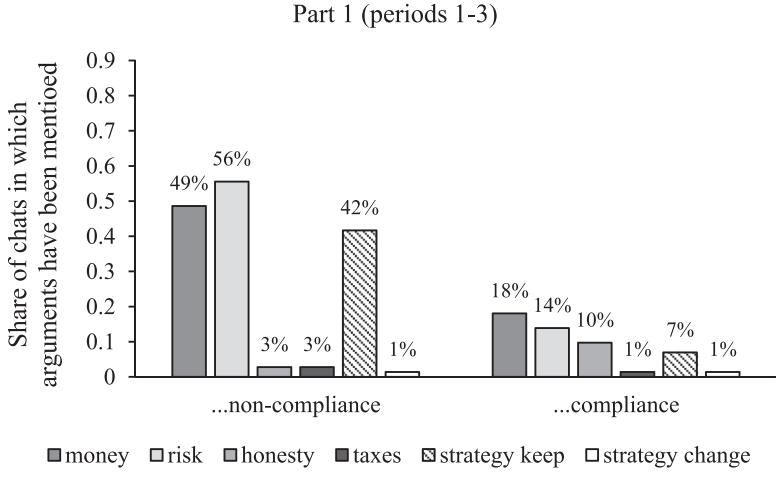


Fig. A7.3. Arguments used in group chats in treatment G–G–G (part 1).

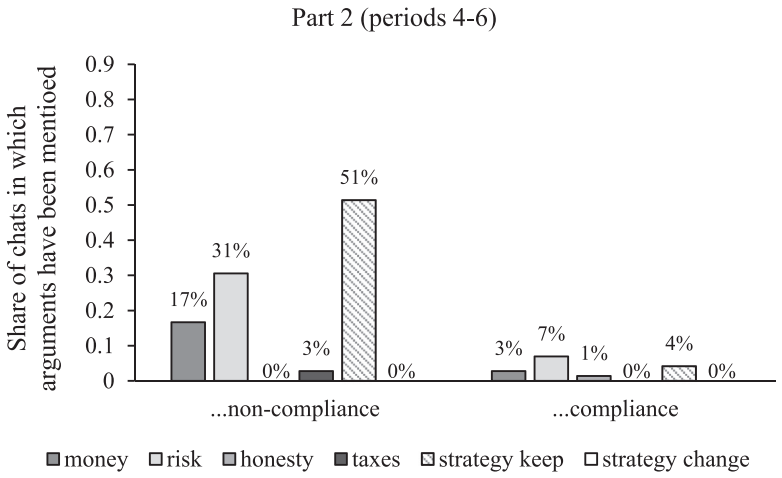


Fig. A7.4. Arguments used in group chats in treatment G–G–G (part 2).

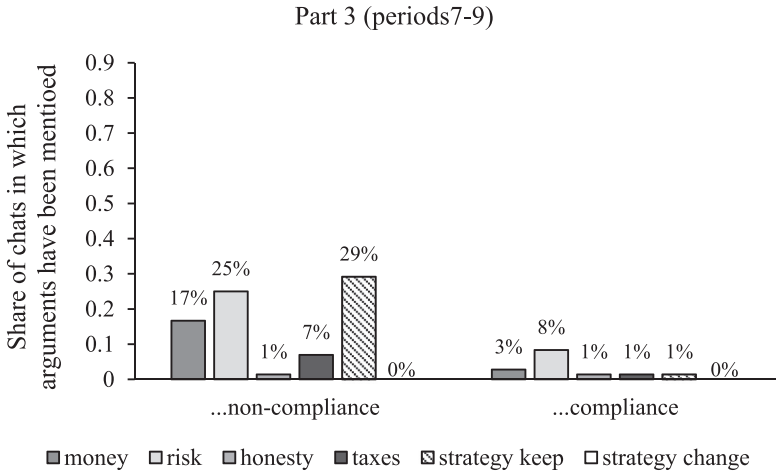


Fig. A7.5. Arguments used in group chats in treatment G–G–G (part 3).

References

Abeler, J., Nosenzo, D., Raymond, C, 2019. Preferences for truth-telling. Econometrica 87, 1115–1153.
Alm, J., 2012. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. Int. Tax Publ. Finance 19, 54–77.
Alm, J., Sanchez, I., De Juan, A., 1995. Economic and noneconomic factors in tax compliance. Kyklos 48, 3–18.
Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. J. Econ. Lit. 36, 818–860.
Azar, O.H., Shira, Y., Bar-Eli, M., 2013. Do customers return excessive change in a restaurant? A field experiment on dishonesty. J. Econ. Behav. Organ. 93, 219–226.
Bäker, A., Mechtel, M., 2015. Peer Settings Induce Cheating on Task Performance IAAEU Discussion Paper Series in Economics, Institut für Arbeitsrecht und Arbeitsbeziehungen in der Europäischen Union, Trier.
Bär, M., Kempf, A., Ruenzi, S., 2011. Is a team different from the sum of its parts? Evidence from mutual fund managers. Rev. Financ. 15, 359–396.
Baker, R.J., Laury, S.K., Williams, A.W., 2008. Comparing small-group and individual behavior in lottery-choice experiments. South Econ. J. 75, 367–382.
Bénabou, R., Falk, A., and Tirole, J. (2019): Narratives, Imperatives, and Moral Persuasion. Working Paper.
Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96, 1652–1678.
Blinder, A.S., Morgan, J, 2005. Are two heads better than one? Monetary policy by committee. J. Money Credit Bank. 37, 789–811.
Bolton, G.E., Ockenfels, A., Stauf, J., 2015. Social responsibility promotes conservative risk behavior. Eur. Econ. Rev. 74, 109–127.
Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group decisions in the centipede game: are groups more “rational” players? J. Exp. Soc. Psychol. 40, 599–605.
Bornstein, G., Yaniv, I., 1998. Individual and group behavior in the ultimatum game: are groups more “rational” players? Exp. Econ. 1, 101–108.
Charness, G., Jackson, M.O, 2009. The role of responsibility in strategic risk-taking. J. Econ. Behav. Organ. 69, 241–247.
Charness, G., Karni, E., Levin, D, 2007. Individual and group decision making under risk: an experimental study of Bayesian updating and violations of first-order stochastic dominance. J. Risk Uncertain. 35, 129–148.
Charness, G., Sutter, M., 2012. Groups make better self-interested decisions. J. Econ. Perspect. 26, 157–176.
Cheung, S.L., Palan, S., 2012. Two heads are less bubbly than one: team decision-making in an experimental asset market. Exp. Econ. 15, 373–397.
Cialdini, R.B., Goldstein, N.J., 2004. Social influence: compliance and conformity. Annu. Rev. Psychol. 55, 591–621.
Cohen, T.R., Gunia, B.C., Kim-Jun, S.Y., Murnighan, J.K., 2009. Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. J. Exp. Soc. Psychol. 45 (6), 1321–1324.

Conrads, J., Irlenbusch, B., Rilke, R.M., Walkowitz, G., 2013. Lying and team incentives. J. Econ. Psychol. 34, 1–7.
Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. Am. Econ. Rev. 95, 477–509.
Cooper, D.J., Rege, M., 2011. Misery loves company: social regret and social interaction effects in choices under risk and uncertainty. Games Econ. Behav. 73, 91–110.
Dufwenberg, M., Dufwenberg, M.A., 2018. Lies in disguise – a theoretical analysis of cheating. J. Econ. Theory 175, 248–264.
Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M., Torgler, B., 2016. Tax compliance and psychic costs: behavioral experimental evidence using a physiological marker. J. Publ. Econ. 134, 9–18.
Erat, S., Gneezy, U., 2012. White lies. Manag. Sci. 58, 723–733.
Ertac, S., Gurdal, M.Y., 2012. Deciding to decide: gender, leadership and risk-taking in groups. J. Econ. Behav. Organ. 83, 24–30.
Fahr, R., Irlenbusch, B., 2011. Who follows the crowd – groups or individuals? J. Econ. Behav. Organ. 80, 200–209.
Feri, F., Irlenbusch, B., Sutter, M., 2010. Efficiency gains from team-based coordination – large-scale experimental evidence. Am. Econ. Rev. 100, 1892–1912.
Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Fochmann, M., Wolf, N., 2019. Framing and salience effects in tax evasion decisions – an experiment on underreporting and overdeducting. J. Econ. Psychol. 72, 260–277.
Gino, F., Ayal, S., Ariely, D, 2009. Contagion and differentiation in unethical behavior. Psychol. Sci. 20, 393–398.
Gino, F., Ayal, S., Ariely, D, 2013. Self-serving altruism? The lure of unethical actions that benefit others. J. Econ. Behav. Organ. 93, 285–292.
Greiner, B. (2015): The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics.
Harrison, G.W., Lau, M.I., Rutström, E.E., Tarazona-Gómez, M., 2013. Preferences over social risk. Oxf. Econ. Pap. 65, 25–46.
He, H., Martinsson, P., Sutter, M., 2012. Group decision making under risk: an experiment with student couples. Econ. Lett. 117, 691–693.
Hofmann, E., Hoelzl, E., Kirchler, E., 2008. Preconditions of voluntary tax compliance: knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. Zeitschrift für Psychologie/J. Psychol. 216, 209–217.
Holt, C.A., Laury, S., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655.
Isenberg, D.J., 1986. Group polarization: a critical review and meta-analysis. J. Pers. Soc. Psychol. 50, 1141–1151.
Janis, I.L., 1972. Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes. Houghton Mifflin, Oxford, England.
Kastlunger, B., Kirchler, E., Mittone, L., Pitters, J., 2009. Sequences of audits, tax compliance, and taxpaying strategies. J. Econ. Psychol. 30, 405–418.
Kocher, M.G., Pogrebna, G., Sutter, M., 2013. Other-regarding preferences and management styles. J. Econ. Behav. Organ. 88, 109–132.
Kocher, M.G., Schudy, S., Spantig, L., 2018. I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. Manag. Sci. 64, 3995–4008.
Kocher, M.G., Sutter, M., 2005. The decision maker matters: individual versus group behavior in experimental beauty-contest games. Econ. J. 115, 200–223.
Korbel, V., 2017. Do we lie in groups? An experimental evidence. Appl. Econ. Lett. 24 (15), 1107–1111.
Kugler, T., Kausel, E., Kocher, M.G., 2012. Are groups more rational than individuals? A review of interactive decision making in groups. Wiley Interdiscip. Rev.: Cognit. Sci. 3, 471–482.
Lahno, A.M., Serra-Garcia, M., 2015. Peer effects in risk taking: envy or conformity? J. Risk Uncertain. 50, 73–95.
Lohse, T., and Simon, S. (2018): Compliance in Teams – Implications of Joint Decisions and Shared Consequences. Working Paper of the Max Planck Institute for Tax Law and Public Finance No. 2018-3.
Maciejovsky, B., Kirchler, E., Schwarzenberger, H., 2007. Misperception of chance and loss repair: on the dynamics of tax compliance. J. Econ. Psychol. 28, 678–691.
Masclet, D., Colombier, N., Denant-Boemont, L., Lohéac, Y., 2009. Group and individual risk preferences: a lottery-choice experiment with self-employed and salaried workers. J. Econ. Behav. Organ. 70, 470–484.
Mazar, N., Aggarwal, P., 2011. Greasing the palm: can collectivism promote bribery? Psychol. Sci. 22 (7), 843–848.
Mittone, L., 2006. Dynamic behaviour in tax evasion: an experimental approach. J. Socio-Econ. 35, 813–835.
Moffatt, P.G., 2015. Experimetrics: Econometrics for Experimental Economics. Macmillan International Higher Education.
Mühlheußer, G., Roider, A., Wallmeier, N., 2015. Gender differences in honesty: groups versus individuals. Econ. Lett. 128, 25–29.
Pahlke, J., Strasser, S., Vieider, F.M., 2015. Responsibility effects in decision making under risk. J. Risk Uncertain. 51, 125–146.
Rau, H.A., 2015. The disposition effect in team investment decisions: experimental evidence. J. Bank Financ. 61, 272–282.
Reynolds, D.B., Joseph, J., Sherwood, R., 2009. Risky shift versus cautious shift: determining differences in risk taking between private and public management decision-making. J. Bus. Econ. Res. 7, 63–78.
Rockenbach, B., Sadrieh, A., Mathauschek, B., 2007. Teams take the better risks. J. Econ. Behav. Organ. 63, 412–422.
Schweitzer, M.E., Hsee, C.K., 2002. Stretching the truth: elastic justification and motivated communication of uncertain information. J. Risk Uncertain. 25, 185–201.
Shupp, R.S., Williams, A.W., 2008. Risk preference differentials of small groups and individuals. Econ. J. 118, 258–283.
Sutter, M., 2005. Are four heads better than two? An experimental beauty-contest game with teams of different size. Econ. Lett. 88, 41–46.
Sutter, M., 2007. Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. Econ. Lett. 97, 128–132.
Sutter, M., 2009. Deception through telling the truth? Experimental evidence from individuals and teams. Econ. J. 119, 47–60.
Torgler, B., 2002. Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. J. Econ. Surv. 16, 657–683.
Weisel, O., Shalvi, S., 2015. The collaborative roots of corruption. Proc. Natl. Acad. Sci. U.S.A. 112, 10651–10656.
Whyte, G., 1993. Escalating commitment in individual and group decision making: a prospect theory approach. Organ. Behav. Hum. Decis. Process. 54, 430–455.
Wiltermuth, S.S., 2011. Cheating more when the spoils are split. Organ. Behav. Hum. Decis. Process. 115, 157–168.