Arbeitskreis Quantitative Steuerlehre
Quantitative Research in Taxation – Discussion Papers

Martin Fochmann, Frank Hechtner, Erich Kirchler, Peter N.C. Mohr

When happy people make society unhappy: How incidental emotions affect compliance behavior

arqus Discussion Paper No. 237

March 2019

www.arqus.info
ISSN 1861-8944

When happy people make society unhappy: How incidental emotions affect compliance behavior

Martin Fochmann[a], Frank Hechtner[b], Erich Kirchler[c], Peter N.C. Mohr[d]

March 2019

Abstract

Emotions have a strong impact on our everyday life, including our mental health, sleep pattern, overall well-being, and judgment and decision making. Our paper is the first study to show that incidental emotions, i.e., emotions not related to the actual choice problem, influence the compliance behavior of individuals. In particular, we provide evidence that individuals have a lower willingness to comply with social norms after being primed with positive incidental emotions compared with aversive emotions. This result is replicated in a second study. As an extension to our first study, we add a neutral condition as a control. Willingness to comply in this condition ranges between the other two conditions. Importantly, this finding indicates that the valence of an emotion but not its arousal drives the influence on compliance behavior. Furthermore, we show that priming with incidental emotions is only effective if individuals are – at least to some extent – emotionally sensitive.

Keywords

Compliance behavior, emotions, cheating, tax evasion, norms, experimental economics

JEL-Classification

C91, D91, H26

[a] Prof. Dr. Martin Fochmann, University of Cologne, Albertus-Magnus-Place, D-50923 Köln, Germany, fochmann@wiso.uni-koeln.de.

[b] Prof. Dr. Frank Hechtner, Technische Universität Kaiserslautern, Gottlieb-Daimler-Straße 42, D-67663 Kaiserslautern, Germany, hechtner@wiwi.uni-kl.de.

[c] Prof. Dr. Erich Kirchler, University of Vienna, Universitaetsstrasse 7, A-1010, Vienna, Austria, erich.kirchler@univie.ac.at.

[d] Prof. Dr. Peter N.C. Mohr, WZB Berlin Social Science Center, Reichpietschufer 50, D-10785 Berlin, Germany, peter.mohr@wzb.eu; Freie Universität Berlin, Garystraße 21, D-14195 Berlin, Germany, peter.mohr@fu-berlin.de.

1 Introduction

Organizations and their managers should act in a socially responsible, ethically clean way that is compliant with social norms (e.g., the law). However, a variety of scandals in the past have damaged this picture. If immoral behavior is disclosed, then negative consequences such as penalty payments, compensation claims, or loss of reputation can result for the organization and for the managers involved. A recent example is Volkswagen's diesel emissions scandal. Volkswagen had to pay enormous penalty fees in the US and in Germany and was faced with a loss of reputation, and the market reacted with a sharp drop in the company's share price and a massive reduction in diesel car demand. The managers involved in scandals face the risk of being held accountable for their actions, which leads to serious consequences. They are fired, and if they face legal proceedings, then large fines or imprisonment can be the consequence. Similar negative consequences have resulted for organizations and their managers involved in corruption (e.g., Walmart, Siemens, Halliburton, KBR), financial accounting fraud (e.g., Enron, Worldcom, Tyco), tax fraud (Paradise Papers, Panama Papers), and insurance fraud (e.g., Hospital Corporation of America, TAP Pharmaceuticals).

A great number of studies argue that social norms encourage individuals to be compliant (e.g., Torgler, 2002; Alm, 2012). This is true even when it is impossible for a lie to be detected (Abeler et al., 2014). It is argued that the moral costs associated with noncompliant behavior reduce the utility of being noncompliant and therefore motivate compliance (Kirchler, 2007; Traxler, 2010). The literature suggests that integral emotions (i.e., emotions related to the actual choice problem), such as guilt or shame felt when violating a social norm, are important drivers of compliance behavior. Along this line, Dulleck et al. (2016) relate social norm violation and the moral costs of noncompliance to aversive emotions and psychic stress.

We contribute to the literature by providing the first paper to analyze how compliance behavior is influenced by incidental emotions (i.e., emotions not related to the actual choice problem). Assume a manager of a car company who is deciding whether to use illegal software to decrease the emissions of cars during tests. Usually, this decision happens in the context of other events that take place before or during the processes of making this choice. The weather might be good or bad, the manager might have been reading good news or bad news in the newspaper, or she might have had meetings with positive or negative outcomes beforehand. All of these events might induce emotions that could interact with the aversive emotions that the manager feels as a result of a possible social norm violation.

In our first study, we use a standardized set of pictures widely applied in psychological research to prime our participants with either positive or aversive emotions. As hypothesized, we are able to show that incidental emotions influence compliance decisions. In particular, compliance is lower after positive incidental priming than after aversive incidental priming. This result is replicated in a second study in which we vary some parameters to test the robustness of our finding. For example, we use pictures from another open-access online database. As an extension to study 1, we also add a neutral condition as a control. Willingness to comply in this condition ranges between the other two conditions. This provides evidence that the valence of affective response (positive or aversive) to our incidental priming and not affective arousal (which is lowest in the neutral condition) drives our results. Furthermore, we show that priming with incidental emotions have a strong impact on compliance behavior for individuals who are emotionally sensitive, but not for individuals who are not.

2 Literature and Hypothesis

2.1 Integral Emotions and Compliance Behavior

The influence of incidental emotions on compliance behavior has not been analyzed to date. However, there is evidence that integral emotions affect compliance. One specific domain of compliance is tax compliance. For example, Erard and Feinstein (1994) propose a theoretical model for the role of "moral sentiments" (i.e., emotions) in tax compliance decisions. Following their approach, two emotions play a central role, shame and guilt, which both arise from the contemplation of imagined actions. In this context, shame is related to the imagination of being caught and punished; can be viewed as an anticipated emotion that can be processed rationally, just as any other consequence of a decision (e.g., monetary fine). In contrast, guilt can be felt independent of detection; it can be viewed as an affective reaction to an intended violation of a social norm, and it arises while making a choice. In the tax compliance context, the social norm might be to declare one's taxes truthfully. If a taxpayer evades taxes, then her behavior is in conflict with this norm, and violating the norm creates an aversive emotional reaction.

Both aversive emotions, shame and guilt, are integral emotions and are strongly related to the concept of the moral costs of tax evasion discussed in the tax literature (e.g., Fortin et al., 2007; Traxler, 2010). Such studies provide robust evidence that tax compliance increases if the moral costs associated with tax evasion increase. Kirchler (2007) suggests that anticipated emotions become "a cost factor in evaluating one's likely advantages and disadvantages of tax

evasion" (p. 64). Therefore, it can be expected that a higher level of aversive emotion associated with potential noncompliance (e.g., higher level of guilt) also increases compliance.

To empirically test this relationship, Dulleck et al. (2016) build on the idea of Erard and Feinstein (1994) and propose a model that integrates "psychic stress", which basically refers to what Erard and Feinstein (1994) defined as guilt. To measure the aversive emotional reaction in tax compliance decisions, they use heart rate variability (HRV), which can be viewed as a correlate of these reactions that may arise from imagining social norm violations. The authors observe a positive correlation between HRV and tax compliance, suggesting that larger aversive emotional reactions to possible social norm violations do in fact lead to more tax compliance, which provides empirical evidence that aversive emotions influence compliance decisions in the predicted direction. Further evidence is provided by Grasmick and Bursik (1990), who argue that aversive emotional reactions associated with tax evasion have a substantial influence on tax compliance decisions. Bosco and Mittone (1997) and Blaufus et al. (2017) show that as an integral emotion, shame/guilt is an effective deterrent of tax evasion, at least in the short run.

In summary, although only a few studies have thus far investigated the role of emotions, integral emotions might be important drivers of compliance behavior and be associated with the moral costs of noncompliance. In particular, an intended violation of a social norm might create an aversive emotional reaction while making a compliance decision, and consequently, it might reduce the tendency to be noncompliant.

2.2 Incidental Emotions and Economic Decision Making

The recent economics literature provides some evidence that incidental emotions impact economic decision making. For example, Capra (2004) induces different incidental emotions through memory elicitation (i.e., participants are asked to recall and write about a sad or happy past event) and through the experience of success/failure during the experiment (i.e., participants are confronted with either difficultly or easily answerable test questions). Her results suggest that participants in a good emotional state are more altruistic in dictator games than participants in a bad emotional state. Kirchsteiger et al. (2006) use funny or sad movies to induce incidental emotions and observe that aversive emotions foster reciprocity, whereas positive emotions foster generosity in a gift-exchange game. Drouvelis and Grosskopf (2016) also use movies and find that angry individuals contribute less to a public good than do happy individuals, resulting in welfare losses. They also show that angry individuals punish more harshly than happy individuals.

Incidental emotions also play an important role in the context of risk taking. Au et al. (2003), for example, find that financial market traders differ in their trading behavior when they are in a good or bad incidental emotional state, which is elicited by music. Traders in a good emotional state show overconfidence and inferior performance, whereas traders in a bad state display more accurate decisions and perform more conservatively. Along this line, Chou et al. (2007) compare risk-taking behavior after priming individuals with positive, aversive, or neutral incidental emotions. The participants in their study show more risk taking after happy priming compared with sad priming.

Some studies also aim to relate incidental emotions to the parameters of specific models for decisions under risk. Rottenstreich and Hsee (2001), for example, hypothesize that the extent of probability weighting, as proposed by prospect theory, depends on the "affective richness" of potential outcomes. In line with their hypothesis, the authors find that "affect-rich" outcomes (e.g., receiving an electric shock or a kiss) are associated with more pronounced probability weighting than are less "affect-rich" outcomes (e.g., money). Schulreich et al. (2014) extend this line of research by showing that individuals choose the riskier lottery significantly more often after being primed with happy music-evoked incidental emotions compared with sad incidental priming and after listening to random tones. The observed difference can be attributed to changes in probability weighting.

In summary, there is robust evidence that incidental emotions can have a substantial impact on social behavior and risk taking. The results from the literature suggest that positive (vs. negative) incidental emotions foster prosocial behavior and the willingness to take risk.

2.3 Hypothesis

Compliance decisions are often operationalized as decisions under risk (Torgler, 2002). Noncompliance usually bears the risk of being penalized with a certain probability. How the tradeoff between obtaining a certain outcome when being compliant and taking the risk of being noncompliant for potentially higher outcomes is solved is, among others, a question of the decision maker's willingness to take risk. In particular, a higher willingness to take risk is associated with lower compliance. Given the evidence from the risk-taking literature that the willingness to take risk is higher after priming with positive incidental emotions, one would expect a lower compliance level after priming with positive incidental emotions compared with aversive incidental emotions.

Furthermore, incidental emotions might interact with integral emotions (Västfjäll et al., 2016). In particular, the effect of an aversive integral emotion resulting from an intended norm

violation might be (1) reduced by inducing a positive incidental emotion and (2) amplified by inducing an aversive incidental emotion. This effect would be in line with the so-called additivity-of-affect hypothesis, which suggests that congruent valences of incidental and integral emotions are added and that incongruent valences cancel each other out (Neumann et al., 2001; Västfjäll et al., 2016). It would also be in line with the dual-emotion concept, which suggests that the effect of an integral emotion can be reduced by inducing a counteracting incidental emotion (Lerner et al., 2015). As the aversive integral emotion might reduce the tendency to be noncompliant, inducing a positive incidental emotion might reduce this positive effect on compliance behavior. In contrast, inducing a negative incidental emotion will amplify (or at least retain) this positive effect. Consequently, we would expect a lower compliance level after positive incidental priming than after aversive incidental priming. In summary, we formulate the following research hypothesis:

Hypothesis 1: *Compliance is lower after positive incidental priming than after aversive incidental priming.*

3 Study 1

3.1 Experimental Design and Sample

Experimental task. To investigate whether incidental emotions influence compliance decisions, we use a tax compliance setting because it mimics several important characteristics of compliance decisions for individuals and organizations. First, a tax compliance decision implies two important dimensions of interest: the honesty dimension, as individuals behave dishonestly when evading taxes (i.e., violating the social norm of paying taxes), and the risk dimension, as tax evasion is disclosed with a certain probability and negative monetary consequences result. Second, the tax compliance literature provides us with a standard and well established framework for our analysis that uses the observed compliance level as an indicator of immoral behavior (Allingham and Sandmo, 1972, Torgler, 2002). Third, the applied tax compliance framework is a nonstrategic task and ensures that strategic uncertainty over others' behavior does not influence one's own behavior.

Our experiment consists of nine periods. In each period, the participants receive a (pretax) income, and their task is to declare the income. A tax of 25% is levied on the declared income. After the participants have submitted their tax return, they are audited with a certain probability. If a participant gets audited and her declared income is lower than her true income, then she has to pay a fine that is twice the amount of the evaded taxes. The implication is that in the case of

a detected tax evasion, the participant has to repay the evaded taxes plus additional penalty costs of 100% of the evaded taxes. The payoff of each period is consequently calculated by the pretax income minus the declared tax minus a potential fine. The participants are not informed about the audit outcome after each period to prevent the audit outcome of one period from influencing tax compliance behavior in subsequent periods.

At the end of the experiment, the computer randomly decides which of the nine periods is paid out. The participants are then informed about the audit outcome and all monetary consequences in the chosen period. The payoff of that period plus a show-up fee of 4 € is paid in cash at the end of the experiment. In our experiment, we use lab-points as the unit of currency, where 1 lab-point exactly corresponds to 0.01 €.

Treatments. We use a between-subjects design with two treatments (positive and aversive emotion treatment) to study how incidental emotions influence tax compliance. To induce incidental emotions, we use pictures from the International Affective Picture System (IAPS; Lang et al., 2008). This database consists of a standardized set of pictures for studying emotions and has been widely applied in psychological research. All pictures are rated and classified into four different categories (e.g., disgust and fear for aversive pictures and amusement and contentment for positive pictures). Pictures that cannot be assigned to one category are classified as undifferentiated.

Aversive pictures are used in our aversive emotion treatment and positive pictures in our positive emotion treatment. All pictures are taken from the undifferentiated category, which has the advantage of having very similar ratings in all four categories. The ratings of the chosen pictures do not differ between genders. We select pictures with mean categorical ratings above two to ensure that the pictures have a relatively high potential to induce incidental emotions.

In each period, a participant faces one picture. As we have a total of nine periods, we use nine different positive and nine different aversive pictures. A picture is displayed only once to a participant. The presentation of a picture starts 10 s before the choice screen is displayed, and the picture remains on the screen until a choice is made, after which a blank screen is shown for 6 s. To exclude the possibility that the participants are always confronted with the same economic environment in our periods, we vary the level of pretax income and audit probability.[1] On the choice screen, the participants are informed about the current pretax income and probability level. The order of the presented picture, pretax income and audit probability combinations used is determined randomly for each participant to control for order effects.

[1] The pretax income is 800, 1200 or 1600 lab-points, and the audit probability is 0%, 33.33% or 66.67%.

Questionnaires. Before the actual experiment starts, the participants are asked to answer a questionnaire that collects sociodemographic data (e.g., age, gender) and information on individual risk attitude. Furthermore, we use the Positive and Negative Affect Schedule (PANAS) with two 10-item scales to measure current positive and negative affect. After the experiment, the participants are asked to rate each picture shown with respect to valence and arousal. Then, the PANAS is presented again to control for how aversive and positive pictures influence affect. Additionally, we collect individual data on tax experience, tax knowledge, tax morale, and income.

Sample. The experiment was conducted at the computerized experimental laboratory of the University of Cologne (CLER) in 2017. The experimental software was programmed and run with the z-Tree software (Fischbacher, 2007), and the participants were recruited with ORSEE (Greiner, 2015). A total of 365 participants (51.5% females) took part in the experiment and earned, on average, 13.90 € (mean duration: 1 hour). A total of 182 participants were randomly assigned to our positive and 183 to our aversive emotion treatment.

The translated instructions (A1) and more information on the pictures used (A2) and our participant pool (A3) can be found in our online appendix.

3.2 Manipulation Checks

Valence rating. To check whether our emotion manipulation was successful, the participants were asked to rate the valence level for each picture presented after all compliance decisions were completed. We applied the following question: "How do you perceive the presented picture?" (9-point Likert scale from 1 = "negative" to 9 = "positive"). The mean (median) valence level for the pictures shown is 7.33 (8.00) in the positive emotion treatment and 2.06 (1.00) in the aversive emotion treatment (see Figure 1); the difference is significant (Mann-Whitney U-test, $p < 0.001$, two-tailed).

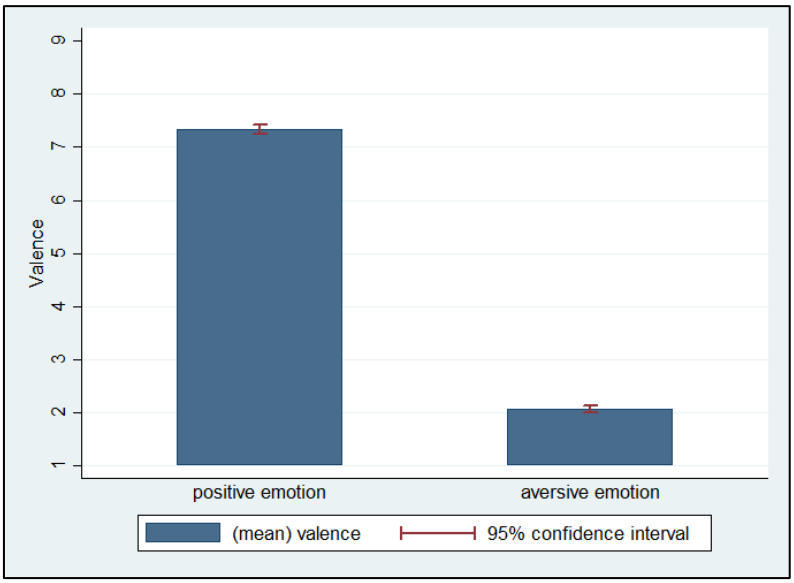


Figure 1: Valence ratings of pictures

PANAS. We used the PANAS to check whether the pictures shown have different effects on current positive and negative affect. For this purpose, we asked our participants to complete the PANAS before and after the main experiment (i.e., before and after the participants were confronted with the pictures). For each type of affect, ten items are used, and the mean rating for all ten items measures the current affect. The higher the level is, the higher the corresponding current affect. Before the experiment, the mean rating for positive (negative) affect is 2.75 (1.62) in the positive emotion treatment and 2.81 (1.67) in the aversive emotion treatment (see Table 1); the differences between treatments are not significant (Mann-Whitney U-test, both p-values above 0.1, two-tailed). As a result, the participants in the two treatments do not differ in their current positive and negative affect before the experiment started. After the experiment, however, we observe a significantly higher positive affect in the positive than in the aversive emotion treatment ($p = 0.026$, two-tailed) and a significantly higher negative affect in the aversive than in the positive emotion treatment ($p < 0.001$, two-tailed).

In summary, these results provide clear evidence that our manipulation was successful.

Table 1: Positive and Negative Affect Schedule (PANAS) test results

		PANAS before experiment		PANAS after experiment		PANAS difference	
		positive emotion treatment	aversive emotion treatment	positive emotion treatment	aversive emotion treatment	positive emotion treatment	aversive emotion treatment
positive affect	mean	2.75	2.81	2.57	2.39	-0.18	-0.43
	MWU test	$p = 0.150$		$p = 0.026$		$p < 0.001$	
negative affect	mean	1.62	1.67	1.48	1.79	-0.14	0.12
	MWU test	$p = 0.617$		$p < 0.001$		$p < 0.001$	

Note: This table shows the results of the PANAS test with two 10-item scales for measuring current positive and negative affect. The higher the level is, the higher the corresponding current affect.

3.3 Results: Compliance Behavior

The main result of our first study is that compliance is higher in the aversive emotion treatment than in the positive emotion treatment. In particular, income is reported truthfully in 44.2% of all decisions in the aversive emotion treatment compared to 40.8% in the positive emotion treatment. The difference between the two treatments is significant (χ^2 test, $p = 0.047$, two-tailed), which supports our first hypothesis.

To corroborate these descriptive and nonparametric results, we use logistic regressions (see Table 2). To take into account that subjects face repeated decision situations, we run logistic regressions with robust standard errors (models 1 and 2) as well as logistic panel regressions with random effects where the subject's identity number is the cross-sectional variable (models 3 and 4). In all models, the dependent variable is the binary variable "compliant behavior", which takes the value of one if income is reported truthfully (0 otherwise). Moreover, we consider a vector of experiment-specific (level of audit probability and pretax income) and participant-specific (gender, tax morale, risk attitude, etc.) variables. However, our variable of interest is an indicator variable for the treatment variation. The positive emotion treatment serves as the default. Thus, the coefficient of our aversive emotion treatment dummy measures the difference between the aversive and positive emotion treatment. In models 1 and 3, we regress on our aversive emotion treatment dummy and on our participant-specific variables. We additionally include the experiment-specific variables in models 2 and 4.

In all four models, the coefficient of the aversive emotion treatment dummy is positive and significant. In models 1 and 2 with robust standard errors at least at the 5%-level and in models 3 and 4 with logistic panel regressions at the 10%-level. Consequently, the compliance level is higher in our aversive emotion condition than in our positive emotion condition. As a result,

we find support for our first research hypothesis. Furthermore, in line with the literature on tax compliance, we observe that a higher audit probability increases tax compliance (Torgler, 2002), and we observe that a higher pretax income decreases tax compliance (Grundmann and Lambsdorff, 2017). We also find the following significant correlations in all four models: participants are less compliant the greater their level of risk seeking and the more complex they perceived the tax-related decisions in our experiment to be and they are more compliant the greater their level of tax morale.

Table 2: Logistic regressions with compliant behavior as the dependent variable

	model 1	model 2	model 3	model 4
aversive emotion treatment	0.18**	0.24***	0.21*	0.38*
	(0.07)	(0.09)	(0.13)	(0.23)
level of audit probability		4.81***		7.26***
		(0.18)		(0.29)
pretax income		-0.00***		-0.00***
		(0.00)		(0.00)
age	0.02**	0.02***	0.02	0.03
	(0.01)	(0.01)	(0.01)	(0.02)
risk attitude	-0.12***	-0.17***	-0.14***	-0.25***
	(0.02)	(0.02)	(0.03)	(0.05)
female	0.10	0.13	0.12	0.20
	(0.08)	(0.09)	(0.14)	(0.24)
tax experience	-0.00	-0.00	-0.01	-0.02
	(0.08)	(0.09)	(0.14)	(0.25)
tax knowledge	-0.01	-0.01	-0.00	0.01
	(0.10)	(0.11)	(0.17)	(0.30)
tax morale	0.09***	0.12***	0.10***	0.18***
	(0.02)	(0.02)	(0.03)	(0.05)
bachelor	-0.08	-0.12	-0.11	-0.21
	(0.08)	(0.09)	(0.14)	(0.25)
economics	-0.03	-0.05	-0.03	-0.05
	(0.08)	(0.09)	(0.14)	(0.25)
decision complexity	-0.04**	-0.06***	-0.06*	-0.10*
	(0.02)	(0.02)	(0.03)	(0.05)
monthly income	-0.00*	-0.00**	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
constant	-0.63***	-1.80***	-0.74*	-2.70***
	(0.23)	(0.32)	(0.41)	(0.76)
observations	3,276	3,276	3,276	3,276
no. of clusters	-	-	364	364
pseudo R^2	0.0260	0.2486		

Note: In this table, the results of logistic regressions are presented with our binary variable "compliant behavior" (which takes the value of one if income is reported truthfully and 0 otherwise) as the dependent variable (regression coefficients, standard errors in parentheses). The positive emotion treatment serves as the default. To take into account that subjects face repeated decision situations, we run logistic regressions with robust standard errors (models 1 and 2) as well as logistic panel regressions with random effects where the subject's identity number is the cross-sectional variable (models 3 and 4). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

3.4 Additional Explorative Analysis: Arousal Rating

Affective reactions are usually measured on two dimensions: valence (i.e., the degree of positive or negative affective response that a picture evokes) and arousal (i.e., the intensity of the affective response that a picture evokes). Our experimental manipulation was designed to reach differences in valence between our conditions. And indeed, we showed in section 3.2 that our manipulation was successful since valence levels differ highly and significantly between both treatments.

However, one might also argue that the second dimension arousal could have influenced behavior. For example, Jahedi et al. (2017) study whether the arousal level of visual stimuli impact economic decision making. They found almost no difference in decision making between images with a low and a high arousal level for arithmetic tasks, impatience, and risk taking in the domain of losses. However, they observed a weak effect for risk taking in the domain of gains (choosing between a sure payoff and a lottery with two equal outcomes). In particular, participants revealed a higher willingness to take risk in case of the high arousal images.

To control for potential arousal effects, our participants were asked to rate the arousal level for each picture presented after all compliance decisions were completed. We applied the following question: “How exciting do you find the presented picture?” (9-point Likert scale from 1 = “little exciting” to 9 = “extremely exciting”). The mean (median) arousal level for the pictures shown is 4.16 (4.00) in the positive emotion treatment and 4.76 (5.00) in the aversive emotion treatment. The arousal difference between both treatments is rather small, but differs significantly in statistical terms (Mann-Whitney U-test, $p < 0.001$, two-tailed).

3.5 Discussion of Study 1

The main finding of study 1 is that priming participants with incidental emotions influence compliance behavior. We observe that compliance is lower after positive incidental priming than after aversive incidental priming. The results of our study also suggest that pictures seems to be a useful instrument to prime participants with positive and aversive incidental emotions. In fact, our pictures induced a significantly higher positive current affect in the positive than in the aversive emotion treatment and a significantly higher negative current affect in the aversive than in the positive emotion treatment (PANAS).

However, we do not only find differences in valence ratings between our conditions but also between arousal ratings. Taking these results into account, we still (moderately) argue that the compliance differences observed between the positive and aversive emotion treatments are

driven by the different positive and negative emotional responses (valence dimension) for two reasons. First, the valence difference between both treatments of 5.27 (= 7.33 – 2.06) is much more meaningful than the arousal difference of 0.60 (= 4.76 – 4.16). Second, if we follow the results of Jahedi et al. (2017) we would expect a higher willingness for the risky noncompliance alternative in the aversive emotion treatment with the higher arousal level. However, we observed exactly the opposite result. Nevertheless, to ensure that the lower compliance level in the positive emotion treatment is driven by the higher valence level and not by the lower arousal level in this treatment (compared to the negative emotion treatment), we introduce a third condition in our second study. This control condition enables us to disentangle the main driver.

4 Study 2

The motivation for our second study is twofold: replication and extension. First, there is a large and ongoing discussion on a replication crisis in different disciplines – including social sciences, psychology and economics. For instance, Camerer et al. (2018) evaluate the replicability of 21 social sciences experiments published in Nature and Science. They found significant effects that point in the same direction as the original studies for 13 studies (62%). Although this result is slightly better than observed in other disciplines (see, for example, Open Science Collaboration 2015 for publications in psychology journals with a "replication success" of 36%), it suggests that replications are also suitable in economics. To replicate the findings of our first study, we conduct a second study in which we slightly vary some parameters in our experiment. However, the compliance decisions made by our participants in the experiment and the approach to use pictures for inducing incidental emotions is not changed.

Second, as an extension to study 1, we add a new control condition to test whether our observed treatment effect is driven by pictures' valence or arousal rating differences. This control condition is called neutral condition as pictures used are rated between the pictures used in the positive and aversive emotion conditions at the valence dimension. However, to disentangle whether the valence or arousal dimension drives our treatment effect, arousal level of pictures used in the control condition is lower than the arousal level of pictures used in the other two conditions.

4.1 Hypotheses

Our second study aims at testing our first hypothesis again. Nothing changes in this regard. Thus, we expect to find that compliance is lower after positive incidental priming than after aversive incidental priming.

Furthermore, we use the added neutral condition to disentangle whether the valence or the arousal dimension drive our results. With respect to the valence dimension, picture ratings in the neutral condition are between the picture ratings in the positive and aversive emotion conditions. In the first study, we observed a higher valence level and a lower compliance level in the positive emotion condition than in the aversive emotion condition. If the valence dimension drives the compliance difference, we would therefore expect that compliance level in the neutral condition is between the levels in the other two conditions. This leads us to hypothesis 2A:

Hypothesis 2A (Valence): *Compliance is higher after neutral incidental priming than after positive incidental priming, but lower than after aversive incidental priming.*

Usually pictures with a relatively high emotional status (aversive or positive) are rated at a higher arousal level than pictures with a relatively neutral emotional status (Kurdi et al. 2017). Therefore, the pictures used in our neutral condition are rated at a lower arousal level than the pictures used in the positive or aversive emotion condition. In our first study, we observed a lower arousal level and a lower compliance level in the positive emotion condition than in the aversive emotion condition. If the arousal dimension drives the compliance difference, we would therefore expect to observe the lowest compliance level in the neutral condition. This leads to the following hypothesis:

Hypothesis 2B (Arousal): *Compliance is lower after neutral incidental priming than after positive or aversive incidental priming.*

4.2 Experimental Design and Sample

Experimental design. We use the same experimental task and setup as in our first study. In the following, we will only highlight the differences. The main difference is that we take pictures from another database to induce incidental emotions. We do this for two reasons. First, we want to replicate our findings with other and more up-to-date images. Second, we want to take pictures with reliable ratings at the valence and arousal dimensions to test our hypotheses 2A and 2B. Unfortunately, IAPS pictures are not rated in this regard. We now take pictures

from the Open Affective Standardized Image Set (OASIS) database introduced by Kurdi et al. (2017). This open-access online database consists of a standardized set of 900 color images. All pictures are rated with respect to valence and arousal. Ratings were obtained in a study with over 800 participants and are highly reliable.

We selected pictures dependent on the valence levels reported in Kurdi et al. (2017). We took pictures with the highest valence levels for the positive emotion treatment and with the lowest valence levels for the aversive emotion treatment. For the neutral emotion treatment, we selected pictures with valence levels exactly between the pictures selected for the positive and aversive emotion treatments. The ratings of the chosen pictures do not differ between genders. All pictures shown and more information on the pictures used can be found in our online appendix A2.

We vary the level of pretax income and audit probability again to change the economic environment in our periods slightly.[2] The order of the presented picture, pretax income and audit probability used is also determined randomly again to control for order effects. Whereas the corresponding audit probability is shown on the choice screen in each period, the pretax income is displayed on a sheet of paper located at the desk of each participant. Before the experiment starts, participants are confronted with two training periods to familiarize with the decision situation.

Questionnaires. We use the same questionnaire design as in study 1. As we expect that the influence of incidental priming might depend on subject's emotional sensitivity, we include an additional questionnaire before the actual experiment starts. We use the 10-item emotional sensitivity scale proposed by Nock et al. (2008) (e.g., "I tend to get emotional very easily"). A factor analysis of the translated version of the scale (in German) resulted as the original scale in only one factor and in a similar reliability (Cronbach's alpha = 0.89).

Sample. The experiment was conducted at the computerized experimental laboratory of the University of Cologne (CLER) in 2018. The experimental software was programmed and run with the z-Tree software (Fischbacher, 2007), and the participants were recruited with ORSEE (Greiner, 2015). A total of 297 participants (56.2% females) took part in the experiment and earned, on average, 13.16 € (mean duration: 1 hour). A total of 102 participants were randomly assigned to our positive condition, 96 to our neutral condition, and 99 to our aversive condition. More information on our participant pool can be found in our online appendix A3.

[2] The pretax income is 800, 900, 1000, …, or 1600 ECU, and the audit probability is 20%, 30%, or 40%.

4.3 Manipulation Checks

Valence and arousal ratings. The mean (median) valence level for the pictures shown is 7.42 (8.00) in the positive, 4.85 (5.00) in the neutral, and 2.04 (1.00) in the aversive emotion treatment (see Figure 2); differences between all treatments are significant (Mann-Whitney U-test, all p-values below 0.001, two-tailed). The mean (median) arousal level for the pictures shown is 4.50 (5.00) in the positive, 3.23 (3.00) in the neutral, and 4.84 (5.00) in the aversive emotion treatment; differences between all treatments are significant (Mann-Whitney U-test, all p-values below 0.01, two-tailed).

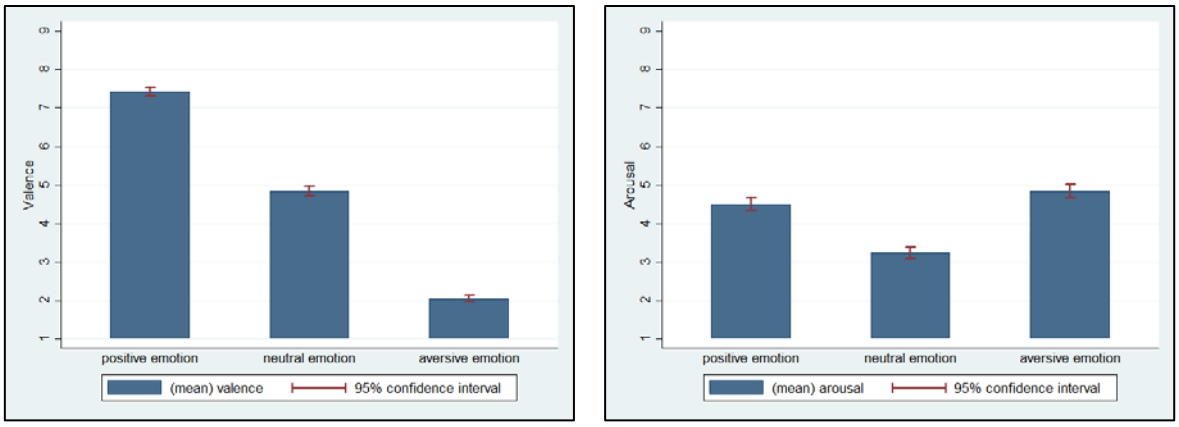


Figure 2: Valence and arousal ratings of pictures in study 2

PANAS. As in the first study, mean ratings for positive and negative affect are very similar across treatments before experiment (see Table 3). After the experiment, however, we observe a significantly higher positive affect in the positive than in the aversive emotion treatment (Mann-Whitney U-test, $p = 0.008$, two-tailed) and a significantly higher negative affect in the aversive than in the positive emotion treatment ($p = 0.004$, two-tailed). Compared to the neutral emotion treatment, we find a significantly higher positive affect in the positive emotion treatment and a significantly higher negative affect in the aversive emotion treatment after the experiment.

In summary, these results provide clear evidence that our manipulation was successful. It is noteworthy that we observe very similar valence, arousal, and PANAS ratings for the positive and aversive emotion treatments in study 1 and 2. Furthermore, our valence and arousal ratings in all three treatments are very similar to the corresponding ratings observed by Kurdi et al. (2017) (see online appendix A2).

Table 3: Positive and Negative Affect Schedule (PANAS) test results (study 2)

	PANAS before experiment			PANAS after experiment			PANAS difference		
	positive	neutral	aversive	positive	neutral	aversive	positive	neutral	aversive
positive affect	2.87	2.67	2.77	2.57	2.37	2.31	-0.30	-0.29	-0.46
MWU tests									
positive vs. aversive		p = 0.150			p = 0.008			p = 0.004	
positive vs. neutral		p = 0.025			p = 0.073			p = 0.722	
neutral vs. aversive		p = 0.517			p = 0.498			p = 0.01	
negative affect	1.60	1.62	1.49	1.50	1.56	1.73	-0.10	-0.07	0.24
MWU tests									
positive vs. aversive		p = 0.174			p = 0.004			p < 0.001	
positive vs. neutral		p = 0.847			p = 0.765			p = 0.939	
neutral vs. aversive		p = 0.291			p = 0.016			p < 0.001	

Note: This table shows the results of the PANAS test for study 2 with two 10-item scales for measuring current positive and negative affect. The higher the level is, the higher the corresponding current affect.

4.4 Results: Compliance Behavior

The main result of our second study is that we are able to replicate the findings of our first study. Income is reported truthfully in 43.1% of all decisions in the positive emotion treatment compared to 52.7% in the aversive emotion treatment. In our neutral emotion treatment, the share lies between the other two treatments and amounts to 47.9% (see Figure 3). The differences between all treatments are significant (χ^2 test, $p < 0.05$ for all comparisons, two-tailed). Consequently, we find support for our hypothesis 1 again and provide evidence for hypothesis 2A. Hypothesis 2B must be rejected.

We use the same logistic regressions and specifications as in our study 1 to corroborate these descriptive and nonparametric results (see Table 4). However, we now also introduce a dummy for our neutral emotion treatment. The positive emotion treatment serves as the default again. Therefore, each coefficient or our aversive and neutral emotion treatment dummy measures the respective difference to the positive emotion treatment. We use a Wald test to check whether both coefficients of the aversive and neutral dummies differ significantly (p-values reported in the last row of table).

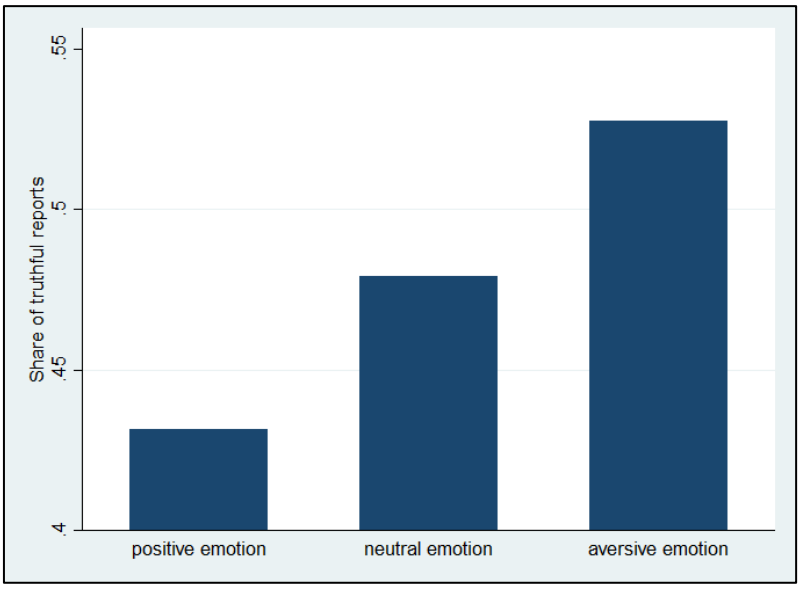


Figure 3: Share of truthful reports in study 2

In line with the results of the first study, the coefficient of the aversive emotion treatment dummy is positive and significant in all four models. In models 1 and 2 with robust standard errors at the 1%-level and in models 3 and 4 with logistic panel regressions at the 10%-level. Thus, we again find support for our research hypothesis. With respect to the neutral condition, we find mixed results. In line with our nonparametric results, the coefficient of the neutral emotion treatment is positive and significant (at the 5%-level) in models 1 and 2. However, in our logistic panel regressions, we do not find a significant effect (although still positive). The difference between aversive and neutral emotion treatment is not significant in any model. Regarding our control variables, we find a positive effect of the audit probability level and age on compliance and a negative effect of the pretax income level in the experiment, risk attitude, decision complexity, economics major, and available monthly income.

Table 4: Logistic regressions with compliant behavior as the dependent variable (study 2)

	model 1	model 2	model 3	model 4
aversive emotion treatment	0.33***	0.35***	0.76*	0.97*
	(0.10)	(0.10)	(0.45)	(0.58)
neutral emotion treatment	0.24**	0.26**	0.56	0.74
	(0.10)	(0.11)	(0.46)	(0.59)
level of audit probability		6.45***		15.38***
		(0.54)		(1.01)
pretax income		-0.00**		-0.00***
		(0.00)		(0.00)
age	0.06***	0.06***	0.12***	0.16***
	(0.01)	(0.01)	(0.05)	(0.06)
risk attitude	-0.17***	-0.18***	-0.37***	-0.49***
	(0.02)	(0.02)	(0.09)	(0.11)
female	0.22**	0.23**	0.39	0.50
	(0.09)	(0.09)	(0.40)	(0.51)
tax experience	-0.09	-0.10	0.04	0.06
	(0.09)	(0.10)	(0.41)	(0.52)
tax knowledge	0.03	0.03	0.03	0.02
	(0.12)	(0.13)	(0.56)	(0.71)
tax morale	0.03	0.03	0.08	0.11
	(0.02)	(0.02)	(0.10)	(0.12)
bachelor	0.26***	0.28***	0.57	0.74
	(0.09)	(0.10)	(0.42)	(0.53)
economics	-0.48***	-0.52***	-0.98**	-1.27**
	(0.09)	(0.09)	(0.41)	(0.52)
decision complexity	-0.11***	-0.12***	-0.21**	-0.27**
	(0.02)	(0.02)	(0.09)	(0.11)
monthly income	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
constant	-0.43	-1.91***	-0.99	-4.70**
	(0.34)	(0.43)	(0.15)	(2.02)
observations	2,655	2,655	2,655	2,655
no. of clusters	-	-	295	295
pseudo R^2	0.0915	0.1346		
Wald test: aversive = neutral treatment	p = 0.391	p = 0.375	p = 0.672	p = 0.691

Note: In this table, the results of logistic regressions are presented for our study 2 with our binary variable "compliant behavior" (which takes the value of one if income is reported truthfully and 0 otherwise) as the dependent variable (regression coefficients, standard errors in parentheses). The positive emotion treatment serves as the default. To take into account that subjects face repeated decision situations, we run logistic regressions with robust standard errors (models 1 and 2) as well as logistic panel regressions with random effects where the subject's identity number is the cross-sectional variable (models 3 and 4). *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

4.5 Additional Explorative Analysis: Emotional Sensitivity

In the following explorative analysis, we examine how the effect of incidental emotions on compliance behavior depends on participant's emotional sensitivity. Our conjecture is that priming with incidental emotions can only be effective if individuals are emotionally sensitive. As outlined in section 4.2, we use the 10-item emotional sensitivity scale proposed by Nock et al. (2008) with answers given on a 5-point Likert scale. For each participant, we calculate the mean over all 10 items for our emotional sensitivity measure. The observed median (mean) emotional sensitivity level in our sample is 2.76 (2.80). Please note that our treatments do not differ significantly in the participants' emotional sensitivity level.

We divided our sample into three subsamples almost equal in size to analyze the compliance behavior of participants with a low, medium, and high emotional sensitivity level (see Figure 4). We observe a large treatment effect in the high and medium emotional sensitivity groups. For the high (medium) group, income is reported truthfully only in 44.4% (42.3%) of the cases in the positive emotion treatment, whereas a value of 59.9% (54.4%) is observed in the aversive emotion treatment. In both subsamples, the treatment difference is significant (χ^2 test, both p-values below 0.01, two-tailed). In contrast, we do not find significant differences in the low emotional sensitivity group (42.3% vs. 43.1%). Therefore, we can conclude that priming with incidental emotions has a strong impact on compliance behavior especially for individuals who are meaningfully emotional sensitive. However, participants with a relatively low emotional sensitivity level do not reveal a substantial behavioral response to our priming with positive and aversive incidental emotions.

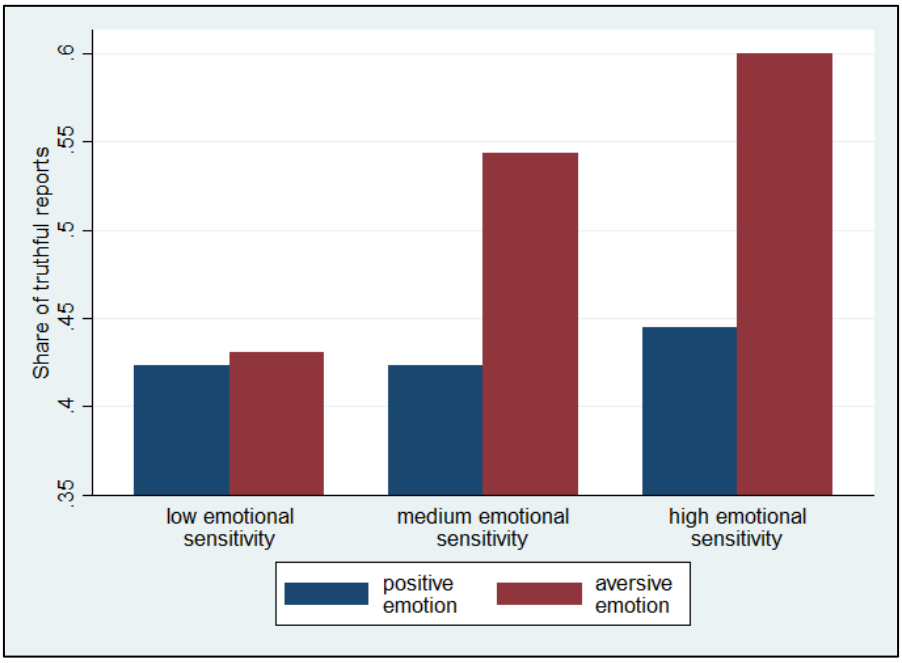


Figure 4: Share of truthful reports in study 2 for our emotional sensitivity groups

4.6 Discussion of Study 2

The main results of our study 2 are threefold. First, we are able to replicate the main finding of our first study that compliance is lower after positive incidental priming than after aversive incidental priming. Second, in our neutral emotion condition as a control, we observe that compliance level lies between the compliance levels of our positive and aversive emotion conditions. This provides evidence that the degree of affective response (positive or aversive) to our incidental priming (valence dimension) drives this finding and not the intensity of affective response that the pictures evoke (arousal dimension). Furthermore, we show that priming with incidental emotions triggers a strong behavioral response for individuals who are emotionally sensitive, but not for individuals who are not.

5 General Discussion and Conclusions

Emotions have a strong impact on our everyday life, including our mental health, sleep pattern, overall well-being, and judgment and decision making (Lerner et al., 2015). The present paper is the first study to show that incidental emotions also influence how individuals comply with the law. In particular, we provide evidence that compliance is lower after positive incidental priming than after aversive incidental priming. We also show that this effect is strongest for individuals who are – at least to some extent – emotionally sensitive.

Our results contribute to the general compliance literature. Although only a few studies to date have investigated the role of emotions, the literature provides robust evidence that integral emotions are important drivers of compliance behavior. We complement the existing literature by showing that compliance behavior is also influenced by induced emotions that are not related to the actual choice problem. For example, our results suggest that the aversive emotions that managers feel as a result of a possible social norm violation when behaving in a noncompliant manner might be lower in a positive than in a negative emotional state. Consequently, the positive effect of such aversive emotions on preventing the managers from behaving in a noncompliant manner is also lowered, and an increase in noncompliance behavior can be expected.

Our findings also contribute to the literature on the influences of emotions on economic decision making, which has only started to study the interactions effects between integral and incidental emotions. Our results are partly in line with the so-called additivity-of-affect hypothesis, which suggests that congruent valences of incidental and integral emotions are added and that incongruent valences cancel each other out (Neumann et al., 2001; Västfjäll et

al., 2016). Priming with positive incidental emotions might have canceled out the effect of aversive reactions to possible social norm violations. This view is also in line with the dual-emotion concept, which suggests that the effect of an integral emotion can be reduced by inducing a counteracting incidental emotion (Lerner et al., 2015). Furthermore, we contribute to the literature on emotions and economic decision making by inducing positive and aversive incidental emotions by means of pictures. To date, different emotional states have been induced by movies (Kirchsteiger et al., 2006; Drouvelis and Grosskopf, 2016; Chou et al., 2007), music (Au et al., 2003; Schulreich et al., 2014) or an imagination/memory task (Capra, 2004; Rottenstreich and Hsee, 2001).

In addition, our findings might have important implications for managerial practice. Managers and other decision makers in organizations might be influenced by incidental emotions when making compliance decisions. Especially in the case of positive incidental emotions, this influence might lead to choices with unintended consequences. To avoid these choices, organizations might want to establish tools for emotion regulation that managers and other decision makers can use before making compliance decisions. Further research should investigate how such tools can be implemented effectively.

Our findings also have implications for society because noncompliance behavior within organizations might cause negative externalities and harm social welfare in many ways. The most prominent recent example is Volkswagen's diesel emissions scandal with potential driving bans for diesel car owners and sharp resale price drops. Noncompliance behavior, however, is an even more general issue. For example, tax noncompliance lowers the tax revenue collected, and the corresponding provision of public goods, corruption and bribery harms the efficiency of resource allocation. Furthermore, financial accounting fraud distorts the allocation of investments.

References

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. Journal of Public Economics 113, 96–104.

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

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

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

Au, K., Chan, F., Wang, D., and Vertinsky, I. (2003). Mood in foreign exchange trading: Cognitive processes and performance. Organizational Behavior and Human Decision Processes 91, 322–338.

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

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

Capra, M. C. (2004). Mood-driven behavior in strategic interactions. American Economic Review 94, 367–372.

Chou, K. L., Lee, T. M. C., and Ho, A. H. Y. (2007). Does mood state change risk-taking tendency in older adults? Psychology and Aging 22, 310–318.

Drouvelis, M. and Grosskopf, B. (2016). The effects of induced emotions on pro-social behaviour. Journal of Public Economics 134, 1–8.

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

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

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

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

Grasmick, H. G. and Bursik Jr, R. J. (1990). Conscience, significant others, and rational choice: Extending the deterrence model. Law and society review, 837–861.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association 1, 114–125.

Grundmann, S. and Lambsdorff, J., (2017). How income and tax rates provoke cheating – An experimental investigation of tax morale. Journal of Economic Psychology 63, 27–42.

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

Kirchsteiger, G., Rigotti, L., and Rustichini, A. (2006). Your morals might be your moods. Journal of Economic Behavior and Organization 59, 155–172.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. Annual review of psychology 66, 33.1–33.25.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. Cognition and Emotion 15, 725–747.

Rottenstreich, Y. and Hsee, C. K. (2001). Money, kisses, and electric shocks: on the affective psychology of risk. Psychological Science 12, 185–190.

Schulreich, S., Heussen, Y. G., Gerhardt, H., Mohr, P. N. C., Binkofski, F. C., Koelsch, S., and Heekeren, H. R. (2014). Music-evoked incidental happiness modulates probability weighting during risky lottery choices. Frontiers in Decision Neuroscience.

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

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

Västfjäll, D., Slovic, P., Burns, W. J., Erlandsson, A., Koppel, L., Asutay, E., and Tinghög, G. (2016). The arithmetic of emotion: Integration of incidental and integral affect in judgments and decisions. Frontiers in Psychology 7, 325.

Appendix (Online Appendix, not intended for publication)

A1 Instructions

Appendix A1 includes the translated instructions (from German) of study 1. All participants received the instructions 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 an experiment in which you have the opportunity to earn money and a questionnaire at the beginning and at the end of the study. How much money you earn depends on your decisions in the experiment and on chance. These instructions explain to you how you can use your decisions to influence how much money you earn in this study.

It is important that you understand the instructions. Therefore, please do not be afraid to ask questions. If you have a question, please raise your hand. We will then come to you to answer your question. Please do not ask your question loudly. You can write and make markings on the instructions. Please do not take the instructions home, but return them to us at the end.

The evaluation of the experiments will be anonymous. Under no circumstances will we associate your name with the data collected in the experiments. You will not get to know the identity of the other participants, neither before nor after the experiment. The other participants will also not be aware of your identity. At the end of the experiment, you must sign a receipt to confirm that you received your payment. The receipt is used for accounting purposes only.

We would like to inform you that you are neither allowed to communicate with other participants nor allowed to leave your desk during the experiment. Please switch off your mobile phone and put it in your bag.

The calculator and pen lying in front of you can be used.

For simplification purposes, this experiment does not calculate in euro amounts, but in Lab-points. One Lab-point is exactly 1 Euro cent. That means, 100 Lab-points equal exactly 1 Euro.

At the end of this experimental study, you will be paid individually and in cash. Your total payout consists of your payout of the experiment and your participation fee.

On the following two pages you will find the instructions for the experiment.

Note: In the experiment, we use visual material that can also be negative (such as images of war and violence). At your own request, you can leave the experiment at any time. In this case, you will receive the participation fee as a payout.

A1.2 Instructions of the Experiment

The experiment consists of a total of 9 periods, which are independent of each other. The following instructions apply to each of these 9 periods.

At the end of the experiment, one period will be randomly selected and paid out.

Income

At the beginning of each period you will be given an income. The amount of income ranges from 800 to 1,600 Lab-points.

From period to period, the amount of income can be different and will be displayed to you during each of your decisions.

Tax return

As part of this study, you must complete a fictitious tax return on your earned income that is used to determine a fictitious tax. Specifically, this means that you should declare your income in each period, which was assigned to you in this period. Of the declared income stated by you in the tax return, a tax of 25% will be deducted. The total tax revenue will be used, among others, to finance future research projects at the University of Cologne.

To declare your income, simply determine how much you want to declare from your actual income, with only integer values possible.

The tax payable is 25% of the declared income:

tax = 0,25 x declared income

Audit of tax return

With a certain probability, your tax return is audited.

This probability may vary from period to period and will be displayed to you during each of your decisions. The level of probability lies within the interval 0% to 66.67%.

If there is an audit and the declared income does not coincide with the actual income, you have to pay the unpaid tax. Additionally, a fine is charged at the same amount.

Additional tax payment = unpaid tax

Fine = unpaid tax

The unpaid tax is:

Unpaid tax = 0.25 x (actual income – declared income)

Period profit and total payout from the experiment

Your period profit which is relevant for your payout is calculated in each period as follows:

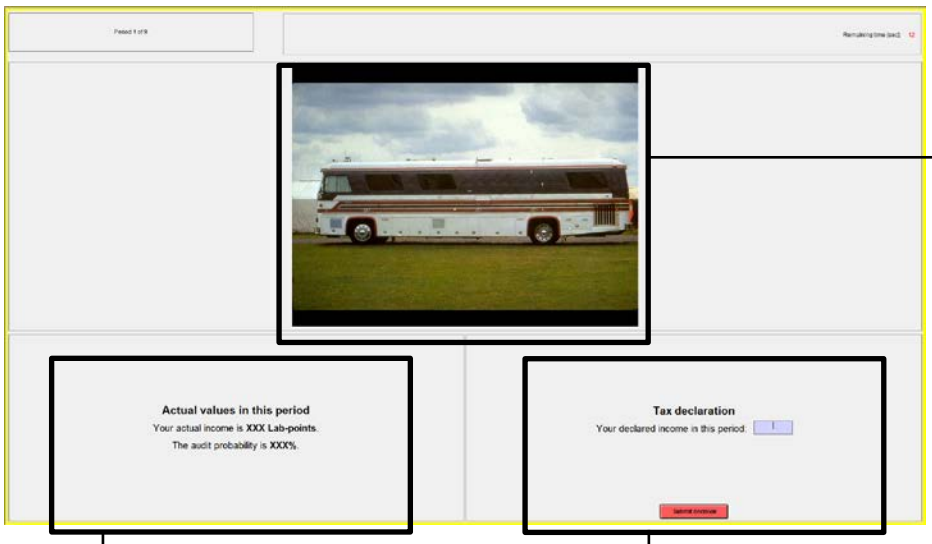
Period profit = actual income – payable tax
– potential additional tax payment – potential fine

After making decisions in all 9 periods, at the end of the experiment, the computer program randomly selects a period and notifies you on the screen. The period profit, which results from this period, will be converted into Euros and added to your participation fee of 4 €. The resulting total payout will be paid to you in cash following the experiment.

Please note that you will not be informed after a period if there has been an audit of your tax return during this period. However, at the end of the experiment, you will be notified if there has been an audit of your tax return in the randomly selected period.

Screen during the decision

While you are making a decision in a period, you will see various information displayed on your screen. The following example serves as an illustration:



At the top of the screen you will see an image in each period. The picture has no relation to your decisions. Just imagine watching TV or reading a magazine while making the decision.

In the left half of the bottom of the screen, you will see in each period the actual income allocated to you during that period and the probability that your tax return will be audited during that period.

In the right half of the bottom of the screen you make your decision.

Final information

Before and after the experiment, you will be asked to answer a few questions on your computer. Answering these questions is not relevant for your payout.

A2 Pictures Used

Study 1

In our first study, we use pictures from the IAPS database (Lang et al. 2008). Table A2.1 presents the IAPS database picture numbers for each treatment. The picture shown in the instructions is also from the IAPS database (no. 7140), but it is a neutral picture, thus categorized as neither positive nor negative. Please notice that we are allowed to use the pictures in our study, but we are not allowed to display these pictures in our paper.

Furthermore, Table A2.1 shows the mean valence and arousal ratings of our participants for each picture. For the valence dimension, we applied the following question: "How do you perceive the presented picture?" (9-point Likert scale from 1 = "negative" to 9 = "positive"). For the arousal dimension, we applied the following question: "How exciting do you find the presented picture?" (9-point Likert scale from 1 = "little exciting" to 9 = "extremely exciting"). Standardized values (minimum value is set to 0 and maximum value is set to 1) are presented in brackets.

Table A2.1: Pictures used in study 1

Picture no.	IAPS database picture no.		Valence rating		Arousal rating	
	positive	aversive	positive	aversive	positive	aversive
1	1710	6312	8.28 (0.91)	1.59 (0.07)	3.76 (0.34)	4.92 (0.49)
2	4641	6350	6.99 (0.75)	1.32 (0.04)	4.53 (0.44)	5.65 (0.58)
3	1610	9280	7.14 (0.77)	3.36 (0.29)	3.51 (0.31)	2.98 (0.25)
4	2057	3500	7.92 (0.86)	1.65 (0.08)	3.96 (0.37)	5.31 (0.54)
5	1460	6020	7.50 (0.81)	1.93 (0.12)	3.88 (0.36)	4.21 (0.40)
6	5623	6300	6.70 (0.71)	2.38 (0.17)	6.23 (0.65)	4.33 (0.42)
7	4601	6230	6.65 (0.71)	2.07 (0.13)	4.59 (0.45)	5.77 (0.60)
8	1440	6940	7.68 (0.83)	2.07 (0.13)	3.63 (0.33)	5.21 (0.53)
9	2501	6830	7.14 (0.77)	2.21 (0.15)	3.34 (0.29)	4.49 (0.44)
1 - 9			**7.33 (0.79)**	**2.06 (0.13)**	**4.16 (0.39)**	**4.76 (0.47)**

Study 2

In our second study, we use pictures from the OASIS database (Kurdi et al. 2017). Table A2.2 displays the selected pictures and the corresponding image file names in the open-access online database.

Table A2.3 shows for each picture the corresponding mean valence and arousal ratings observed in the study by Kurdi et al. (2017) – KLB in the following – and the ratings of the participants in our second study. Please notice that KLB used a 7-point Likert scale whereas we used a 9-point Likert scale for both dimensions. Standardized values (minimum value is set to 0 and maximum value is set to 1) are presented in brackets for comparison purposes.

Table A2.2: Pictures used in study 2
(below each image: image file name and theme in Kurdi et al. 2017)

No.	Positive emotion treatment	Neutral emotion treatment	Aversive emotion treatment
1			
	Lake 9	Desert 1	Dead bodies 2
2			
	Fireworks 2	Paperclips 1	Animal carccass 5
3			
	Lake 1	Camping 9	Garbage dump 4
4			
	Mother 1	Power lines 1	Fire 5
5			

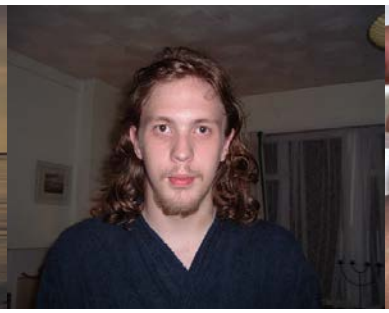
	Galaxy 7	Sleepy pose 4	War 1
6			
	Beach 4	Roofing 3	Explosion 5
7			
	Couple 7	Car 2	Bloody knife 1
8			
	Flowers 8	Gorrila 1	Dog 24
9			
	Bird 1	Neutral face 4	Shot 3

Table A2.3: Pictures used in study 2: valence and arousal ratings

Picture No.		Valence rating			Arousal rating		
		positive	**neutral**	**aversive**	**positive**	**neutral**	**aversive**
1	KLB study	6.41 (0.90)	4.00 (0.50)	1.25 (0.04)	4.11 (0.52)	3.27 (0.38)	4.48 (0.58)
	Our study	7.69 (0.84)	4.66 (0.46)	1.87 (0.11)	4.52 (0.44)	3.25 (0.28)	4.85 (0.48)
2	KLB study	6.28 (0.88)	3.99 (0.50)	1.62 (0.10)	4.98 (0.66)	2.00 (0.17)	4.48 (0.58)
	Our study	7.15 (0.77)	4.88 (0.48)	1.97 (0.12)	4.27 (0.41)	3.44 (0.31)	4.47 (0.43)
3	KLB study	6.26 (0.88)	3.99 (0.50)	1.64 (0.11)	3.97 (0.50)	3.50 (0.42)	3.79 (0.47)
	Our study	7.41 (0.80)	4.94 (0.49)	1.96 (0.12)	4.31 (0.41)	3.30 (0.29)	5.03 (0.50)
4	KLB study	6.11 (0.85)	3.97 (0.50)	1.74 (0.12)	4.09 (0.52)	2.67 (0.28)	5.29 (0.72)
	Our study	7.61 (0.83)	5.04 (0.51)	2.18 (0.15)	4.69 (0.46)	3.05 (0.26)	4.67 (0.46)
5	KLB study	6.07 (0.84)	3.96 (0.49)	1.76 (0.13)	4.50 (0.58)	2.53 (0.26)	5.00 (0.67)
	Our study	7.29 (0.79)	4.84 (0.48)	2.05 (0.13)	4.78 (0.47)	3.00 (0.25)	4.86 (0.48)
6	KLB study	5.93 (0.82)	3.93 (0.49)	1.82 (0.14)	3.63 (0.44)	2.08 (0.18)	5.46 (0.74)
	Our study	7.24 (0.78)	4.62 (0.45)	2.11 (0.14)	4.52 (0.44)	3.24 (0.28)	4.91 (0.49)
7	KLB study	5.92 (0.82)	3.92 (0.49)	1.84 (0.14)	4.25 (0.54)	2.87 (0.31)	4.55 (0.59)
	Our study	7.59 (0.82)	4.76 (0.47)	2.19 (0.15)	4.16 (0.4)	3.34 (0.29)	4.86 (0.48)
8	KLB study	5.91 (0.82)	3.91 (0.48)	1.89 (0.15)	3.7 (0.45)	3.54 (0.42)	4.77 (0.63)
	Our study	7.37 (0.80)	5.04 (0.51)	1.87 (0.11)	4.46 (0.43)	3.30 (0.29)	5.22 (0.53)
9	KLB study	5.91 (0.82)	3.89 (0.48)	1.92 (0.15)	3.01 (0.34)	2.18 (0.20)	4.17 (0.53)
	Our study	7.45 (0.81)	4.88 (0.48)	2.20 (0.15)	4.77 (0.47)	3.15 (0.27)	4.66 (0.46)
1 - 9	**KLB study**	**6.09 (0.85)**	**3.95 (0.49)**	**1.72 (0.12)**	**4.03 (0.50)**	**2.74 (0.29)**	**4.67 (0.61)**
	Our study	**7.42 (0.80)**	**4.85 (0.48)**	**2.05 (0.13)**	**4.50 (0.44)**	**3.23 (0.28)**	**4.84 (0.48)**

A3 Main Characteristics of Participants

Table A3.1 provides an overview of the main characteristics of our participants.

Table A3.1: Main characteristics of participants

Variable	**description**	**study 1 (N = 365)**	**study 2 (N = 297)**
female	female = 1; male = 0	51.5%	56.2%
risk attitude	risk question from the German Socio-Economic Panel (0 to 10; no willingness to take risk = 0; high willingness to take risk = 10)	4.64	4.54
age	in years	25.1	24.1
economics	study with more than one lecture in economics = 1 (0 otherwise)	46.8%	52.9%
bachelor	study with a bachelor's degree = 1 (0 otherwise)	52.6%	59.9%
tax experience	experience with tax returns = 1 (0 otherwise)	43.8%	39.4%
tax knowledge	tax knowledge = 1 (0 otherwise)	19.5%	14.5%
monthly income	in Euro (monthly income after fixed costs)	381.6	379.8
tax morale	0 to 9; low tax morale = 0; high tax morale = 9	6.8	7.2
decision complexity	0 to 10; low perceived decision complexity in experiment = 0; high perceived decision complexity in experiment = 10	1.8	2.4

Impressum:
Arbeitskreis Quantitative Steuerlehre, arqus, e.V.
Vorstand: Prof. Dr. Ralf Maiterth (Vorsitzender),
Prof. Dr. Kay Blaufus, Prof. Dr. Dr. Andreas Löffler
Sitz des Vereins: Berlin

Herausgeber: Kay Blaufus, Jochen Hundsdoerfer, Martin Jacob, Dirk Kiesewetter, Rolf J. König, Lutz Kruschwitz, Andreas Löffler, Ralf Maiterth, Heiko Müller, Jens Müller, Rainer Niemann, Deborah Schanz, Sebastian Schanz, Caren Sureth-Sloane, Corinna Treisch

Kontaktadresse:
Prof. Dr. Caren Sureth-Sloane, Universität Paderborn,
Fakultät für Wirtschaftswissenschaften,
Warburger Str. 100, 33098 Paderborn,
www.arqus.info, Email: info@arqus.info

ISSN 1861-8944