The effects of income windfalls on labor supply and tax compliance: experimental evidence*

Miguel A. Fonseca and Lutfi Rahimi

February 28, 2022

Abstract

We examine the effect of cash windfalls on the supply of effort and tax compliance in a high stakes, real effort laboratory experiment. Income windfalls have a negative effect on effort, which mirrors existing evidence from the field. Tax compliance on labor income is only marginally affected by windfalls. When windfall income is taxable, we find a positive relationship between windfall size and compliance. We find no evidence that taxpayers switch from one type of evasion to the other as a function of whether windfalls are taxable or not. While compliance levels are lowest among the most productive, there are no differential effects of windfalls on compliance along the productivity dimension.

Keywords: Cash windfalls, real effort, tax compliance, experiments.
JEL classification: C91, J20, H24, H26.

*We thank Sebastian Kripfganz and David Reinstein for their valuable comments and suggestions. Financial support from ESRC grant ES/N00762X/1 is gratefully acknowledged. The usual disclaimer applies. Email addresses: Fonseca: m.a.fonseca@exeter.ac.uk; Rahimi: lutfira@gmail.com

1 Introduction

Universal Basic Income (UBI), the unconditional provision of a basic level of income to every citizen in a country, has re-entered the public debate in many western societies. While the original motivation for the adoption of UBI was as an effective means to lift large numbers of people out of poverty (Hum and Simpson, 1993), contemporary advocates for UBI argue that it is also a solution to the technology-driven disruptions to the labor market, such as the increased automation in manufacturing and the emergence of the 'gig economy' (Stern, 2016; van Parijs and Vanderborght, 2017).

The appeal of UBI lies in its conceptual and administrative simplicity: either as a cash transfer or as a negative income tax, its unconditional nature allows the state to do away with costly layers of bureaucracy required by means-tested programs; unlike policies like unemployment insurance, cash windfalls do not suffer from moral hazard (Fabre et al., 2014). The arguments against unconditional cash transfers run along two lines: their high implementation costs (Ortiz et al., 2018) and their potentially detrimental impact on labor supply (Hum and Simpson, 1993). Most of the attention in the literature on UBI and other unconditional cash transfers has been on their effects on labor market outcomes (Hum and Simpson, 1993; Marinescu, 2018).[1]

While many taxpayers' income is reported by their employers, a substantial proportion of the labor force—estimated to be about 30% in the US—are freelancers or work in the 'gig economy' (US Federal Reserve, 2019). Such individuals self-report their income for tax purposes as they are, for all intents and purposes, self-employed.[2] Self-employed taxpayers are more likely to engage in tax non-compliance than employees (Kleven et al., 2011). In fact, the problem of misreporting income was already prevalent in the original 1970s negative income tax experiments, where recipients under-reported income in order to get a larger payout (Robins, 1985; Burtless, 1986; Ashenfelter and Plant, 1990). In short, tax compliance is an important behavioral response one must consider when examining income windfalls. This is particularly so in economies in which freelance work accounts for a large proportion of the labor force, since tax revenue is an important determinant of the sustainability of such a program.

In the present paper, we examine the effect of unconditional income transfers on tax compliance as well as effort supply. We tackle our research question using real-effort laboratory experiments. Laboratory experiments allow for true ceteris paribus comparisons; they also enable us to study the effects of cash transfers on the full distribution of productivity. Importantly, we can directly observe evasion, which is extremely difficult and costly to do in the

[1] Some studies have focused on health and other socio-economic outcomes; see Marinescu (2018) for a review and discussion.

[2] Recent rulings around Uber mean this may change, but at the time of writing this paper, the statement applies.

field, while retaining a measure of external validity, as evasion in the lab correlates strongly with evasion in the field (Dai et al., 2018). That said, the lab is an imperfect approximation to reality, as the decisions we study are made in a highly stylized environment and in a very short span of time. We mitigate this limitation by conducting a high stakes experiment, whereby the average payment to participants for an hour-long session was equal to the average week's salary in the area where we conducted the experiments.

In every round of our experiment, participants had to perform a real-income task which was their source of labor income. They also received a non-labor income windfall that was randomly determined in every round. In the Non-Taxable Windfall (NTW) treatment, the windfall was non-taxable and participants did not have to declare it for tax purposes. In the Taxable Windfall (TW) treatment, the windfall was taxable and had to be declared for tax purposes separately.

For an individual who has reference-dependent preferences (Kahneman and Tversky, 1979; Kőszegi and Rabin, 2006), a cash windfall decreases willingness to work because it puts the individual closer to their income target, and therefore less effort is required to reach it. In addition, by moving the individual closer to their earnings target, a cash windfall may make tax evasion more attractive. The intuition is that individuals with reference-dependent preferences who have an earnings target operate in the loss domain, in which they are risk seeking. Since the gradient of the value function is steeper the closer we are to the reference point, this means that the upside risk from evasion will be more attractive than the downside risk from being caught. This extra utility from evasion may also compensate for the loss of income from lower labor supply. The latter effect is magnified by the fact that labor income is earned in our experiment, which makes it subjectively more valuable to the taxpayer on the margin (Kirchler et al., 2009). We therefore predicted that windfalls would increase non-compliance and reduce effort by participants in our experiment.

We find that real-effort provision declines with windfalls. However, that decline is not a function of the magnitude of the windfall: non-taxable windfalls have a smaller negative effect on effort supply than taxable windfalls. This suggests a tax aversion effect akin to that found by Kessler and Norton (2016), except in our case it operates through tax on a cash windfall. Labor income tax compliance is negatively affected by cash windfalls. When windfall income is taxable, we find a positive relationship between windfall size and windfall income tax compliance. We find no evidence that taxpayers switch from one type of evasion to the other as a function of whether windfalls are taxable or not. While compliance levels are lowest among the most productive participants, there are no differential effects of cash windfalls on compliance along the productivity dimension.

Our paper lies at the intersection of two large experimental literatures: one focusing on labor supply (see Charness and Kuhn, 2011 for an extensive review) and another on tax compliance (see Mascagni, 2018 for an extensive literature review of lab experiments on

compliance, and Alm and Malezieux, 2021 for a recent meta-analysis). By examining supply of effort and tax compliance jointly, this literature is able to explore how labor markets adjust to changes in fiscal parameters. Such estimations are typically difficult to accomplish in the field, especially when tax compliance is involved. Doerrenberg and Duncan (2014) also implement a lab experiment where participants complete a real-effort labor task and declare taxes on their income. They focus on the behavioral impact on compliance and labor supply of changing the tax rate. The authors study the impact on the decision to comply or not, as well as the role of tax rate changes on labor supply. Collins et al. (1992) study the same question, but focusing on the effect of the tax schedule. They compare no taxation to tax schedules that vary in their progressivity. Pántya et al. (2016) compare the effect of flat vs. progressive tax schedules on participants' labor supply and tax compliance. They make this comparison between-subjects, as well as within-subjects. The latter is done by changing the tax schedule mid-experiment. Their experiment differs from ours in that tax receipts were multiplied by 1.5 and the resulting funds were equally re-distributed among participants in the session. Therefore tax paying has a public good provision element that is absent in our design. Our study differs from Pántya et al. (2016) in that we vary the amount of windfall income assigned to participants within a proportional tax system with regards to labor income, while that paper examines the effect of a progressive taxation system in the absence of windfall payments.

Our paper also contributes to the literature on unconditional cash transfers. With a view to resolve the question of whether universal cash windfalls would adversely affect labor supply, a number of UBI experiments, dating back to the 1970s set out to empirically estimate the effect of UBI on labor supply. Most have focused on the poorest, as opposed to a representative sample (Hum and Simpson, 1993; Marinescu, 2018).[3,4] As such, these programs cannot inform us about the consequences windfalls have on the full distribution of labor productivity. However, evidence on the effects of other income windfalls like state lotteries or the Alaska Permanent Fund corroborates the evidence on UBI programs: labor supply is only marginally negatively affected (Robins, 1985; Burtless, 1986; Imbens et al., 2001; Akee et al., 2010; Cesarini et al., 2017; Jones and Marinescu, 2018; Picchio et al., 2018). To the best of our knowledge, there has not been an experimental examination of the effect of unconditional cash transfers on tax compliance.

[3]Most 1970s UBI/NIT experiments had a cut-off rule for participation rule as a function of the poverty line (Hum and Simpson, 1993). The Finnish basic income experiment had a random sample, but it consisted solely of individuals who were unemployed on January 1, 2017. The Alaska Permanent Fund Dividend is the only income windfall program that was universally applied to all eligible citizens. Since it was a policy change, the econometric analysis of its effects has relied on quasi-experimental methods.

[4]Another issue that likely contributed to an overestimation of labor supply effects is selective attrition: Ashenfelter and Plant (1990) show that treated families who received no payment (because they had higher outside earnings) and therefore likely had no labor supply response to the NIT were more likely to leave the experiment, affecting the sample.

2 Experimental Design and Procedures

In this section we outline the experimental hypotheses and describe the experimental design.

2.1 Theoretical Framework and Hypotheses

We employ a prospect theory framework to model the behavior of participants in our experiment. Reference-dependent models have become quite popular to model taxpayer behavior (see among many others, Chang et al., 1987; Schepanski and Kelsey, 1990; Copeland and Cuccia, 2002; Kirchler et al., 2009). The critical assumption in reference-dependent models is the reference point of the decision maker.

We assume that the reference point is the expectation formed by each individual participant on their future experimental earnings. The invitation email to the experiment stated that participants would receive 100 AFN (at the time of the experiment approximately £1, $1.30) for taking part, plus an additional payment that would depend on their decisions. It is therefore plausible that participants expected to earn an additional amount in the same order of magnitude as the show-up fee. The fact that their additional earnings would be of the same magnitude as the show-up fee should have been apparent once they read the instructions to the experiment, which detailed the task, the earnings and the conversion rate from the experimental currency unit to AFN. We could not measure the expectations of each participant on their earnings, but we assume it is positive, reasonably large and heterogeneous.

We now discuss the implications of such reference point preferences on effort in the experiment, as well as tax compliance. Taking expectations about experimental earnings as the reference point (Kőszegi and Rabin, 2006), a participant with prospect theory preferences will be operating in the loss domain. A windfall will move the participant closer to their income target. As such, the participant will have to work less hard to achieve their target. However, the more loss averse an individual is, the greater the marginal utility of each unit of effort expended, and the weaker this effect will be. There is empirical evidence supporting this prediction in different settings: income-targeting by taxi drivers in New York (Camerer et al., 1997) and Singapore (Chou, 2002), as well as the model of loss aversion and labor supply where participants exhibit diminishing sensitivity by working less on days when the wages are high (Goette et al., 2004). Abeler et al. (2011) show that expectations about earnings have a significant impact on effort in a laboratory setting: participants whose income expectations were manipulated to be high worked harder than those whose income expectations were manipulated to be low.

Hypothesis 1. *Effort decreases with provision of windfalls, irrespective of whether the windfall is taxed or untaxed.*

As a corollary from the above, taxed windfalls will cause a smaller change in income than

untaxed windfalls relative to an agent's reference point. Therefore we would expect that the reduction in effort will be larger when windfalls are not taxed.

Hypothesis 2. *The decrease in effort (in absolute level) will be larger under Non-Taxable Windfalls than Taxed Windfalls.*

Note, however, that this model predicts effort levels to decrease once participants reach their income target. A negative relationship between effort and experimental round could be interpreted as participants reaching their target.

We now move to the second dimension of the decision task, tax compliance. Again, assuming prospect theory preferences, we recall that an earnings target means participants operate in the loss domain. A windfall leads participants to move closer the origin (reference point) as their earnings increase. This places them in the steeper part of the utility function in the negative quadrant, where individuals are most risk seeking. Truthful reporting therefore becomes equivalent to a sure loss, while under-reporting is a gamble between a larger loss (if detected) and smaller loss (if undetected).

In principle, the source of earnings has an impact on individual proclivity to give it away as taxes (Arkes and Blumer, 1985; Thaler and Johnson, 1990; Davis et al., 2010). Furthermore, since labor income is earned, that increases the utility from this source of income (Kirchler et al., 2009), making the value function steeper. As such, we would expect lower reported labor earnings when compared to that of windfall income earnings. Separating the tax reporting stages for income windfalls and earnings from the real-effort task allows us to observe any differences. We therefore state our third hypothesis:

Hypothesis 3. *Tax compliance falls with windfall income.*

The fact that we are able to study tax compliance in a country with very weak, incipient institutions is an interesting opportunity to understand the extent to which non-utilitarian motives matter to tax compliance. The literature has identified a moral dimension to tax compliance. Either because of group affiliation, religiosity, a non-trivial fraction of individuals in society comply with their tax obligations out of a sense of duty (Torgler and Schneider, 2009; Halla, 2012, Dwenger et al., 2016). Cummings et al. (2009, p. 448) define *tax morale* as "the intrinsic motivation to pay taxes arising from the moral obligation to pay taxes as a contribution to society". Tax morale as an individual attitude is typically collected in survey instruments. Evidence has steadily accumulated of its relationship with perceptions of non-compliance at national level (Frey and Torgler, 2007), and the size of the shadow economy (Torgler, 2005b; Alm and Torgler, 2006; Torgler and Schneider 2007, 2009; Torgler, Schaner and Macintyre, 2007; Halla, 2012). Finally, tax morale is positively correlated with tax compliance in the lab (Bosco and Mittone, 1997; Torgler et al., 2007). We take the latter approach and study the correlation of survey responses capturing tax morale with behavioral

non-compliance in the lab. This approach has the drawback of the questionnaire data being elicited after the experiment has taken place. It is possible that individuals may seek to confirm or justify their behavior in the experiment in the way they respond to tax morale questions. We therefore do not make a causal interpretation to this hypothesis.

Hypothesis 4. *Tax compliance is positively correlated with stated tax morale.*

2.2 Experimental Design

Each round of the experiment was divided in three stages. In stage one, that round's income windfall was publicly announced to participants (we used a pre-determined random sequence to ensure comparability across sessions). The windfall amount was 0, 15 or 30 tokens, corresponding to our No, Low and High treatments. The non-labor income windfall was made salient by forcing participants to type the amount on their screens and they could not lie at this stage. We followed a neutral framing referring to the windfall as "variable tokens" in the experiment.

In stage two, participants chose to take part in a real effort task to earn more income or opt out to move on to the next stage of the experiment. The explicit choice was imposed to identify those participants who were not willing to provide effort, rather than to rely on them not engaging with the task itself – the choice to leave stage two was always available to participants during the whole of stage two, to allow participants to proceed with the experiment once they reached their optimal number of sliders before the time limit elapsed. They had two minutes in which to finish the task and earn as much income as they wished. The real effort task used was the slider task developed by Gill and Prowse (2012) in which 48 sliders are scattered across the computer screen. Participants must move each slider from its initial position on the far left to the middle position. Participants received one token for each correctly positioned slider. Charness and Kuhn (2011) point out that hours worked and labor supply are isomorphic and share many characteristics; therefore our real effort task is a good proxy for labor supply, particularly given the high stakes. Furthermore, the task is intrinsically meaningless, which minimize any experimenter demand effects or gift exchange motives for participant effort.[5] The first two stages were repeated 12 times, thus each participant had each windfall four times. To see the decision screens see Appendix D.

In stage three, participants were informed of their total earnings and were asked to report their earnings for tax purposes. Following the literature on tax compliance, in our experiment tax was referred to as "tax" in the experiment (Alm and Malezieux, 2021). In the Taxable Windfall treatment, they were asked to report their real effort and windfall earnings separately. They could report earnings between zero up to their gross income. The fact that participants

[5] We note that it may be impossible to completely rule out experimenter demand effects, even with neutral framing, as documented by Dürsch and Müller (2017).

have to report one single source of non-labor income which comes from the same agency to which taxpayers declare their income may, at first, appear strange. We could have added other sources of non-labor income that are jointly reported with the windfall to better approximate reality. However we felt that the additional complexity did not justify the benefit of greater external validity. In the Non-Taxable Windfall treatment, participants were asked to report earnings from the real effort task only. At the end of each round, participants saw their total earnings and total tax deductions individually. Participants were not told of the earnings of other participants in the room.

After participants completed the 12th round, we collected an incentivized measure of risk aversion based on Gneezy and Potters (1997), a question on tax morale taken from the World Values Survey, and a measure of the big-5 personality scale (Woods and Hampson, 2005). Subjects were then paid individually in cash, on the basis of their average payoff of three randomly drawn rounds plus their show-up fee to avoid a bad-round pay (Charness et al., 2016).

We kept the following parameters constant to match as closely as possible the context of Afghanistan, where the experiment took place: the audit probability was set at 5% and the penalty amount was equal to 25% of the evaded tax if underreporting was discovered plus the unpaid tax liabilities.[6] The tax rate in the experiment was 35% of the taxable income. In short, our experimental design varied within-subjects on the amount of non-labor income: None (0 tokens), Low (15 tokens), and High (30 tokens); it varied between-subjects whether or not non-labor income was taxable.

Each treatment had 12 sessions. The first treatment took place between August and November 2017 and the second treatment was carried out between November and December 2018. We recruited from the same participant pool in different years but the procedures were kept the same. Sessions were held at different times of the day to minimize sample selection issues. Recruitment emails asked participants to take part in a decision-making experiment informing them that they would be paid 100 AFN (approximately £1, $1.30) and that they could earn more depending on their decisions and chance. There was no mention in the email text of labor supply, tax evasion or compliance. The experiment was programmed in z-Tree (Fischbacher, 2007).

The participant pool used for both experiments were part-time students from Rana University, a privately-owned institution in Kabul, Afghanistan. Virtually all of our sample had some form of business and/or work experience. As such, our participants will have faced labor/leisure supply decisions, and had experience with the incipient tax system in Afghanistan.

The relatively low levels of income in Afghanistan allow us to implement a very high-stakes experiment; the weighted average weekly income in Kabul is estimated to be $8.3 and $7.4 for

[6] The probability of getting audited in the experiment at least once was about 46% ($1 - 0.95^{12} = 0.46$).

	TW			NTW		
Windfall level	0	15	30	0	15	30
Effort	14.52	14.23	12.88	15.98	15.45	15.28
	(5.06)	(5.02)	(4.83)	(5.44)	(5.65)	(5.76)
Labor income	0.49	0.47	0.48	0.42	0.43	0.42
tax compliance rate	(0.36)	(0.37)	(0.35)	(0.31)	(0.32)	(0.29)
Non-labor income	0.00	0.44	0.57			
tax compliance rate	(0.00)	(0.32)	(0.41)			
Aggregate	0.49	0.45	0.53	0.42	0.43	0.42
tax compliance rate	(0.36)	(0.30)	(0.36)	(0.31)	(0.32)	(0.29)
	215	215	215	215	215	215

Standard deviations in parentheses.

Table 1: Average Effort and Compliance

2017 and 2018 respectively (Akseer et al., 2017; 2018). Most work in experimental economics focuses on WEIRD (Western, Educated, Intelligent, Rich, and Democratic) samples (Henrich et al., 2010); tax compliance/labor economics experiments are no exception (Charness and Kuhn, 2011; Alm and Malezieux, 2021). We extend the evidence base in this area to a non-Western sample with limited experience with democratic government and a fully-functioning state. The average duration of each session was 70 minutes including payment time. The average payment was around $8, including the show-up fee, corresponding to roughly one week's average income.

3 Results

Our analysis will focus on two outcome variables: effort, measured by the number of sliders solved in a given round; and compliance, measured as the ratio of declared to true income. We will test average treatment effects using average compliance by an individual over the course of the experiment using standard parametric tests.[7] We employ random effects GLS models which take advantage of the panel structure of our experimental data; the unit of analysis is the behavior of an individual at the round level. In the following sections, we will present our main results with support.

3.1 Effort

Result 1. *Windfalls lead to a reduction in labor supply.*

[7]We note that we have a large enough sample that allows us to use t-tests for comparing means; Wilcoxon-Mann-Whitney tests give similar results.

Support. Table 1 summarizes average effort and average compliance conditional on windfall level for both TW and NTW treatments. We note that the average individual effort over the course of the experiment ranged between 4 and 27 with a median of 15. This means that the smallest windfall condition doubled the income of the median participant and more than doubled the income of the least productive participant.

Effort significantly declines with windfall in TW ($\chi^2(2) = 14.431, p < 0.001$, Kruskal-Wallis (KW) test), but not in NTW ($\chi^2(2) = 1.674, p = 0.433$, KW test). Testing pairwise differences in mean effort confirms this result (TW, Wf0 vs. Wf15: $t(428) = 0.60, p = 0.549$; Wf15 vs. Wf30: $t(428) = 2.83, p = 0.005$; Wf0 vs. Wf30: $t(428) = 3.43, p = 0.001$. NTW, Wf0 vs. Wf15: $t(428) = 0.98, p = 0.326$; Wf15 vs. Wf30: $t(428) = 0.31, p = 0.760$; Wf0 vs. Wf30: $t(428) = 1.29, p = 0.199$.)

We complement our analysis of mean effort by using a random effects GLS estimator to estimate average treatment effects on supply of effort. The dependent variable is the effort level of participant i in round t. It takes integer values between 0 and a theoretical maximum of 48 (the number of sliders). However, the maximum value was never reached, avoiding the need to use truncated models. The basic specification estimates average treatment effects through dummies for the different windfalls (Wf15 and Wf30); the zero-windfall condition is the omitted category. We consider an additional specification with a time trend and individual characteristics, including risk tolerance, which captures diminishing marginal utility of income (Gneezy and Potters, 1997)[8], gender, and age.

Table 2 summarizes the results of the estimations. In all models, the omitted treatment is the zero-windfall condition. Model (1) estimates the most parsimonious model that only includes treatment dummies. The coefficients on both Wf15 and Wf30 are negative, although only the latter is significant at standard levels. Further, both coefficients are significantly different from each other ($\chi^2(1) = 123.95, p < 0.001$), suggesting a non-linear relationship between windfalls and labor supply. Model (2) includes a number of personal characteristics that we collected after the experiment concluded, in particular risk aversion, gender and age, as well as a time trend. We find a significant effect of risk aversion, which is consistent with the effect of concavity of the utility function: diminishing marginal utility of money means a smaller effect of a given piece rate on the margin, and therefore lower effort. Importantly, neither the sign nor the magnitude of the treatment effects are affected by the inclusion of individual-level controls.

Model (3) estimates the basic model for the NTW treatment. The coefficients on both Wf15 and Wf30 are negative and significant; however they are not significantly different from each other ($\chi^2(1) = 0.63, p = 0.426$), suggesting a more linear relationship between windfalls and labor supply in this environment. Model (4) extends it by considering the effects of risk

[8]Our risk tolerance variable takes values between 0 and 4, where the latter corresponds to risk neutral/seeking individuals and the remainder to different degrees of risk aversion.

	Taxable Windfall		Non-taxable Windfall		All
DV: Effort	(1)	(2)	(3)	(4)	(5)
Wf15	-0.292	-0.197	-0.527***	-0.532***	-0.527***
	(0.156)	(0.159)	(0.188)	(0.198)	(0.188)
Wf30	-1.634***	-1.586***	-0.695***	-0.698***	-0.695***
	(0.179)	(0.181)	(0.192)	(0.191)	(0.192)
Wf15×TW					0.235
					(0.244)
Wf30×TW					-0.938***
					(0.262)
TW					-1.462***
					(0.507)
Risk tolerance		0.735**		0.528	
		(0.340)		(0.405)	
Age		-0.030		0.119	
		(0.152)		(0.121)	
Male		0.898		0.720	
		(0.660)		(0.751)	
Round		0.190***		0.002	
		(0.030)		(0.021)	
Constant	14.517***	11.148***	15.979***	11.305***	15.979***
	(0.345)	(3.702)	(0.371)	(3.172)	(0.371)
N	2,580	2,580	2,580	2,580	5,160
R^2	0.01	0.05	0.002	0.01	0.03
Wald χ^2	139.7	214.0	15.53	18.67	178.8

Note: Standard errors clustered at the participant level. Wf15 corresponds to Low condition (15 tokens) and Wf30 to High condition treatment (30 tokens). ***, **: $p < 0.01, p < 0.05$, respectively. In all estimations, we reject the null hypothesis of joint equality to zero at the 1% level.

Table 2: Determinants of effort

aversion, gender and age. We again observe a positive effect of risk aversion, although the coefficient is not significant. Neither age nor gender coefficients are significant.

We now look at Hypothesis 2, pertaining to the effect of taxing windfalls on effort.

Result 2. *The reduction in effort resulting from a non-taxable windfall is larger than or equal to the reduction in effort resulting from a taxable windfall.*

Support. Model (5) in Table 2 estimates the effect of windfalls on effort. The coefficients on *Wf15* and *Wf30* are both negative and significant (*Wf15*: $\chi^2(1) = 7.88, p = 0.005$; *Wf30*: $\chi^2(1) = 13.16, p < 0.001$), indicating that non-taxable windfalls lead to a decrease in effort. The coefficients on TW $\times$ *Wf15* and TW $\times$ *Wf30* capture the difference in the effect of windfalls on effort as a function of its tax status. The coefficient on TW $\times$ *Wf15* is non-significant ($\chi^2(1) = 0.93, p = 0.336$), indicating windfalls of 15 tokens have the same effect on effort irrespective of their tax status. In contrast, the coefficient on TW $\times$ *Wf30* is negative and highly significant ($\chi^2(1) = 12.81, p < 0.001$). This means that taxable windfalls have a smaller impact on effort than non-taxable windfalls.

3.2 Tax Compliance

We next move to the analysis of tax compliance.

Result 3. *Windfalls have no or little negative effect on labor income compliance.*

Support. We start by revisiting Table 1. We find no effect of windfalls on average labor income compliance rate either in the TW treatment ($\chi^2(2) = 0.398, p = 0.820$, KW test) or the NTW treatment ($\chi^2(2) = 0.142, p = 0.931$, KW test). However, the decision whether or not to under-report was made after the effort decision, and therefore we should condition our analysis on the effort decision. We regressed the labor income compliance rate (defined as the ratio of declared labor income to effort) on effort and windfall dummies using a random effects Tobit estimator to account for the truncation of compliance at 0 and 1 (see Figures 1, 2 and 3 in Appendix for the distribution of compliance). We considered a separate specification where we control for risk aversion, demographics, as well as proxies for the big-5 personality traits, attitudes towards government, mistrust in government, tax morale, and knowledge of how to evade Afghanistan's tax system. Risk aversion is the fundamental attribute in choice under uncertainty in the expected utility framework, as well as in prospect theory (albeit differently in loss/gain domains). Attitudes and mistrust in government are included in the regression as we are conducting our experiment in a country with limited experience of democracy and in which the government has limited enforcement capacity. Therefore such attitudes may be important determinants of compliance. Knowledge of evasion proxies experience with evasion, which may correlate with willingness to evade. Tax morale captures non-utilitarian motives

for compliance. Finally, Alaheto (2003) and Choo et al. (2016) find correlations between some of the big-5 personality traits and white-collar crime and compliance respectively, so we include it here as an exploratory set of variables.

Model (1) in Table 3 estimates average treatment effects without individual controls. The coefficients on the windfalls are both negative, but not significant (Wf15: $\chi^2(1) = 2.59, p = 0.108$; Wf30: $\chi^2(1) = 0.66, p = 0.415$). Model (2) extends the previous model by considering individual characteristics. The coefficient on effort remains negative and highly significant ($\chi^2(1) = 10.07, p = 0.002$). Both windfall coefficients retain their sign, are not significant (Wf15: $\chi^2(1) = 2.79, p = 0.095$; Wf30: $\chi^2(1) = 0.66, p = 0.414$).

Model (5) estimates the determinants of labor income tax compliance when windfalls are not taxable. The coefficient on Wf15 is very close to zero and non-significant ($\chi^2(1) = 0.04, p = 0.844$); the coefficient on Wf30 is negative but non-significant ($\chi^2(1) = 1.61, p = 0.204$). Model (6) accounts for individual characteristics. The coefficients on Wf15 ($\chi^2(1) = 5.53, p = 0.019$); Wf30 ($\chi^2(1) = 6.50, p = 0.011$) are now negative and significant.

The analysis has focused on the intensive margin of evasion. We now look at the extensive margin: that is, the decision whether or not to evade at all. To this effect, we replace the dependent variable in the previous analysis with an indicator variable that equals one if participant i declared the full amount on round t and zero otherwise. We estimate this model using a random effects probit estimator. Table 4 summarizes the results.

Model (1) summarizes the basic model for the TW treatment. We find negative coefficients on Wf15 ($\chi^2(1) = 7.42, p = 0.006$) and Wf30 ($\chi^2(1) = 3.25, p = 0.072$), though only significant in the former case. Model (2) extends model (1) to account for individual characteristics. The sign, size and significance of the coefficients on Wf15 and Wf30 is unchanged. Model (5) replicates model (1) this time looking at the NTW treatment. The coefficient on Wf15 is essentially zero and non-significant ($\chi^2(1) = 0.00, p = 0.973$), while the coefficient on Wf30 is negative and not significant ($\chi^2(1) = 3.23, p = 0.073$). Model (6) extends model (5) to take into account individual characteristics. The estimated coefficients on Wf15 and Wf30 are now both negative and significant ($\chi^2(1) = 7.86, p = 0.005$; $\chi^2(1) = 9.53, p = 0.002$, respectively).

Result 4. *Windfall size has a positive effect on windfall income compliance.*

Support. Model (3) in Table (4) replicates model (1) this time looking at windfall tax compliance. The coefficient on effort is negative and significant ($\chi^2(1) = 6.29, p = 0.012$). The coefficient on Wf30 is positive and highly significant ($\chi^2(1) = 50.04, p < 0.001$). Model (4) extends the model to include individual characteristics. The coefficients on effort and Wf30 retain their sign, size and significance level.

DV	Taxable Windfall Labor Income Tax Compliance		Taxable Windfall Windfall Income Tax Compliance		Non-taxable Windfall Labor Income Tax Compliance	
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	-0.014***	-0.013***	-0.037***	-0.039***	-0.021***	-0.021***
	(0.004)	(0.004)	(0.013)	(0.013)	(0.002)	(0.002)
Wf15	-0.056	-0.058			0.003	-0.044**
	(0.035)	(0.035)			(0.018)	(0.019)
Wf30	-0.029	-0.029	0.836***	0.826***	-0.022	-0.045**
	(0.035)	(0.035)	(0.104)	(0.104)	(0.018)	(0.018)
Extraversion		-0.033		-0.086		0.003
		(0.031)		(0.055)		(0.011)
Agreeableness		0.005		0.038		-0.001
		(0.030)		(0.053)		(0.014)
Emot. Stab.		0.033		0.101		-0.028***
		(0.029)		(0.053)		(0.010)
Conscient.		0.068**		0.111**		0.022**
		(0.030)		(0.055)		(0.011)
Openness		-0.041		0.002		0.011
		(0.028)		(0.050)		(0.012)
Att to Gov't		-0.027		-0.019		-0.017
		(0.027)		(0.048)		(0.010)
Tax Morale		0.058		0.028		0.027**
		(0.033)		(0.059)		(0.012)
Mistrust Gov't		-0.061**		-0.088		-0.004
		(0.027)		(0.049)		(0.010)
Knows Evasion		-0.019		0.038		0.0001
		(0.026)		(0.047)		(0.010)
Risk Tolerance		-0.159**		-0.293**		0.031
		(0.074)		(0.134)		(0.032)
Age		0.034		-0.083		0.013
		(0.034)		(0.061)		(0.009)
Male		-0.091		-0.061		0.069
		(0.154)		(0.275)		(0.058)
Round		-0.005		0.024		0.016***
		(0.004)		(0.015)		(0.002)
Constant	0.708***	0.524	0.595**	2.551	0.734***	0.139
	(0.100)	(0.873)	(0.231)	(1.571)	(0.042)	(0.286)
N	2,580	2,580	2,580	2,580	2,580	2,580
LL	-1801.2	-1788.0	-1388.3	-1378.3	-1567.4	-1528.3
Wald χ^2	14.5	42.2	77.8	90.0	142.6	42.2

Note: Standard errors clustered at the individual level in parentheses. Wf15 corresponds to Low condition (15 tokens) and Wf30 to High condition treatment (30 tokens). ***, **: $p < 0.01, p < 0.05$ respectively. In all estimations, we reject the null hypothesis of joint equality to zero at the 1% level.

Table 3: Determinants of tax compliance: intensive margin

We now turn to the role of individual-level heterogeneity and how it impacts our treatment effects. We start by examining the role of productivity differences in our real-effort task.

Result 5. *More productive individuals are more likely to evade on both their labor income and windfall income irrespective of how windfalls are taxed.*

Support. We start by looking at the Tobit estimates on Table 3. In Model (1) which pertains to the TW treatment, we find a negative and highly significant coefficient on Effort ($\chi^2(1) = 11.87, p < 0.001$); controlling for individual characteristics does not change the size or significance of the coefficient.

Doing the same estimation for the NTW treatment (Model (5)), we find that the coefficient on Effort is negative and significant ($\chi^2(1) = 141.00, p < 0.001$). Controlling for personal characteristics (Model (6)) does not change the sign or significance ($\chi^2(1) = 149.49, p < 0.001$).

We next look at the effect of productivity on windfall tax compliance, which is examined in Model (3) in Table 3. Again the coefficient on Effort is negative and highly significant ($\chi^2(1) = 8.20, p = 0.004$). Accounting for individual characteristics does not change the sign or significance of the estimated coefficient on effort ($\chi^2(1) = 8.49, p = 0.004$).

We now move to Table 4, which reports on marginal effects from a random effects probit regression examining the decision whether or not to evade. In Model (1) that pertains to the TW treatment, we find a negative coefficient on effort ($\chi^2(1) = 29.24, p < 0.001$); controlling for individual characteristics does not change its sign, magnitude or significance. Model (5) replicates the analysis for the NTW treatment. We find a negative significant coefficient on effort ($\chi^2(1) = 99.72, p < 0.001$) which is virtually unchanged when we account for individual characteristics. Finally, we examine windfall tax compliance on the extensive margin. Model (3) estimates the basic model pertaining windfall income compliance; the coefficient on effort is negative and significant ($\chi^2(1) = 6.29, p = 0.012$) and is unchanged by the inclusion of individual characteristics.

We further explore the role productivity differentials may play in determining tax compliance. Since compliance is a function of effort, it is possible that windfalls may have had a differential effect on the propensity to evade for individuals with higher productivity levels.

Result 6. *We do not find differential effects of windfalls on labor income tax compliance on the basis of productivity.*

Support. We re-estimate models (2) and (5) in Table 3 and interacted the windfall dummy variables with dummies for the three terciles of the ability distribution (as measured by the distribution of average effort at the individual level). Table 8 summarizes the results; we relegate this table to the appendix for the sake of brevity in exposition. We do not find any

difference in propensity to evade as a function of ability, including a robustness check in which ability was measured linearly.

We conclude our analysis of results by looking at the effect of attitudinal measures, personality traits, and behavioral proxies on compliance.

Result 7. *Conscientiousness and tax morale are strongly correlated with compliance.*

Support. We start by examining the intensive margin of tax compliance. Model (2) in Table 3 summarizes the estimated coefficients on the regression of labor income tax compliance in the TW treatment. Conscientiousness ($\chi^2(1) = 5.00, p = 0.025$) and tax morale ($\chi^2(1) = 3.18, p = 0.075$) have a positive coefficient, though the latter is not significant. Mistrust in government has a significant negative coefficient ($\chi^2(1) = 4.89, p = 0.027$), as does risk tolerance ($\chi^2(1) = 4.58, p = 0.032$). In the NTW treatment estimation (model (6)), the coefficients on emotional stability ($\chi^2(1) = 7.48, p = 0.006$); conscientiousness ($\chi^2(1) = 4.15, p = 0.042$), and tax morale are significant ($\chi^2(1) = 5.37, p = 0.021$), but the coefficient on attitudes toward government ($\chi^2(1) = 3.08, p = 0.079$) is not. Conscientiousness is also associated with windfall income compliance (Model (4): $\chi^2(1) = 4.17, p = 0.041$), as is risk tolerance ($\chi^2(1) = 4.74, p = 0.029$).

Looking at the extensive margin estimations, we again find the estimated coefficients on conscientiousness to be positive and significant in all but one estimation (model (2): $\chi^2(1) = 2.91, p = 0.069$; model (4): $\chi^2(1) = 8.07, p = 0.005$; model (6): $\chi^2(1) = 7.40, p = 0.007$); Tax Morale is positive and significant, but only with respect to labor income tax compliance ($\chi^2(1) = 5.78, p = 0.016$). Risk tolerance is also negatively associated with the decision to comply in the TW treatment, but surprisingly positively associated in the NTW treatment.

4 Discussion

Our experiment finds a negative effect of windfalls on effort, which in turn proxies labor supply (Charness and Kuhn, 2011). This is consistent with existing evidence documenting small and short-lived effects in the field (Cesarini et al., 2017; Marinescu, 2018; Picchio et al., 2018). We must caveat our results by highlighting some small difference in average effort between the TW and NTW treatments when windfalls are absent. This could be due to sample selection issues, although we find this unlikely: the samples for both treatments were drawn from the same participant pool and both samples are quite similar in gender, risk tolerance and several other attitudinal profiles (see Table 5 in the Appendix for details). The TW sample is slightly younger, but not to the extent that would suggest a higher degree of proficiency in the task.

Tax aversion or the additional tax-induced effects due to taxes on windfall may explain the differences in NTW and TW. According to this line of reasoning, heuristics, biases in

DV	Taxable Windfall				Non-taxable Windfall	
	Labor Income Tax Compliance		Windfall Income Tax Compliance		Labor Income Tax Compliance	
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	-0.013***	-0.013***	-0.006**	-0.007***	-0.017***	-0.018***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Wf15	-0.035**	-0.035***			0.000	-0.036***
	(0.013)	(0.013)			(0.013)	(0.013)
Wf30	-0.025	-0.025	0.114***	0.112***	-0.024	-0.042***
	(0.014)	(0.014)	(0.016)	(0.016)	(0.013)	(0.013)
Extraversion		-0.010		-0.012		0.001
		(0.010)		(0.009)		(0.005)
Agreeableness		0.000		-0.001		-0.008
		(0.009)		(0.009)		(0.007)
Emot. Stab.		0.010		0.016		-0.014***
		(0.010)		(0.009)		(0.004)
Conscient.		0.017		0.025***		0.013***
		(0.010)		(0.009)		(0.005)
Openness		-0.011		0.007		0.012**
		(0.009)		(0.009)		(0.006)
Att to Gov't		-0.006		0.001		-0.011**
		(0.009)		(0.008)		(0.006)
Tax Morale		0.025**		0.004		0.012**
		(0.011)		(0.011)		(0.006)
Mistrust Gov't		-0.009		-0.014		-0.008
		(0.009)		(0.008)		(0.005)
Knows Evasion		-0.006		0.004		0.002
		(0.008)		(0.009)		(0.004)
Risk Tolerance		-0.038		-0.060***		0.042***
		(0.021)		(0.022)		(0.014)
Age		0.007		-0.016		0.007
		(0.011)		(0.011)		(0.004)
Male		-0.007		0.008		0.057**
		(0.011)		(0.048)		(0.027)
Round		0.002		0.005		0.012***
		(0.002)		(0.003)		(0.002)
N	2,580	2,580	1,720	1,720	2,580	2,580
L Pseudo-L	-751.7	-742.9	-833.9	-823.4	-853.6	-810.3
Wald χ^2	44.3	57.6	55.4	70.4	87.4	151.6

Note: Coefficients are marginal effects. Wf15 corresponds to Low condition (15 tokens) and Wf30 to High condition treatment (30 tokens). Standard errors clustered at the individual level in parentheses. ***, **: $p < 0.01, p < 0.05$ respectively. LPL and Wald statistics is the Log Pseudolikelihood and Wald statistics from the random effects probit regression from which the marginal effects posted in this table are calculated. In all estimations, we reject the null hypothesis of joint equality to zero at the 1% level.

Table 4: Determinants of tax compliance: extensive margin

thinking and propensity to fall prey to saliency effects are the principal drivers (McCaffery and Baron, 2006; Sussman and Olivola, 2011, although Olsen et al., 2019 failed to replicate Sussman and Olivola's findings in the US and the UK). That said, Kessler and Norton (2016) document that individuals travel longer distances or wait longer in queues to avoid paying additional taxes. Should this effect be at work in our experiment, it operates through the windfall income channel.

Our main hypotheses hinge on the assumption of participants forming expectations about their earnings in the experiment, which in turn operate as a reference point for determining effort and compliance (Kirchler et al., 2009). Effectively, therefore, participants operate in the loss domain. Such expectations, should they exist, are heterogeneous and unobservable since we did not manipulated them experimentally. It is possible that for some participants the expectations were low enough that they might operate in the gain domain, either from the start of the experiment, or at some point in the experiment. We are unable to estimate the former possibility; the latter would be observed through a shift in effort and/or evasion over the course of the experiment. We find limited evidence for that hypothesis in our regression analysis.

Our econometric analysis explicitly relied on an exclusion restriction on the estimation equation for effort. In particular we assume that (expected) tax compliance is not a determinant of the effort decision; otherwise, we do not have identification. To check for the validity of our exclusion restriction, we constructed and performed a simultaneity test on the experimental data using a 2SLS approach. We regressed effort supply on the explanatory and control variables while using a class of instrumental variables (IVs) for tax evasion to test whether the estimated coefficient for evasion is important. We used tax morale as the instrument. The literature on tax morale supports the idea that the social norm of compliance or tax morale is an important driver of compliance (Torgler, 2007). To the best of our knowledge, there are no empirical and/or theoretical studies that have linked tax morale to effort decisions. We find no evidence that compliance is a determinant of effort in our experimental data (see Table 7 in the Appendix for details). A Granger causality test confirms our analysis that effort does not Granger-cause tax compliance.

Our hybrid experimental design while benefiting from the advantages of both between- and within-designs, minimizes the demand-effects with the use of a meaningless task (slider puzzles) both to the participants and the experimenters, and making it salient that participant decisions had no financial bearing on the experimenters. To minimize confounds to identification, carry-over effects, biases to specific orders, and referencing or sensitization to changes, we used a pre-determined random sequence for the order of exposure to windfalls. We segmented the decisions (effort supply and tax compliance) with the time it took the real-effort task to lapse and announcing labor earnings to ensure independent evaluation of each windfall treatment. While this may not be very long in absolute terms but proportional

to the real-effort duration, it is significant.

We find a significant negative correlation between effort and compliance, both at the extensive and intensive margins. This suggests that higher ability individuals are more likely to evade and will evade by a greater extent than low ability individuals. This finding mirrors that by Choo et al. (2016) who, like this paper, employed a sample that included people in full-time employment and self-employed. One possible explanation could be a greater sense of entitlement by those participants who were able to solve large numbers of sliders, akin to the psychological process described by Kirchler et al. (2009) related to task difficulty and value. In this case, hard-earned performance may have prompted higher marginal valuation, a steeper loss function, and as a result less compliance. We cannot verify this conjecture, but it is a promising avenue of future research.

In contrast, we find a non-linear effect of windfalls on the extensive margin in that a windfall of 15 tokens decreased the probability of compliance by 3.5 percentage points. A further increase in the windfall by 15 tokens had no impact on compliance in the taxable windfall treatment and a further decrease in the probability of compliance (although that effect was not consistent across specifications). In contrast, the impact of windfalls on the extent of compliance was negative but non-significant in the taxable windfall treatment and negative and significant in the non-taxable windfall – although again, this effect was only significant in some econometric specifications. In contrast, windfall size positively affects the likelihood of compliance and increases the amount of windfall income declared. The reason for the weak, inconsistent effect of windfalls on labor income compliance could be that participants in our experiment allocated windfall income to a separate mental account to labor income. This is certainly a plausible mechanism in the real-world since any UBI payments (or any windfall income from the government) would likely have a different frequency and origin to labor income. This is a very interesting and promising line of inquiry going forward.

We find that key attitudinal variables have a consistent correlation with tax compliance, whether labor income or windfall income. In particular, our proxy for tax morale exhibited a strong correlation in the predicted direction. We note that our measure of tax morale was collected in a post-experimental questionnaire; we cannot rule out that participants responded to that question in a manner that justifies their decisions in the experiment itself. Therefore we cannot make causal claims about this relationship. With that caveat in mind, our evidence therefore extends the evidence base for tax morale to a society in which there is very limited experience of a fully functioning state, whose capacity for enforcement of tax compliance is quite limited.

Likewise, our conscientiousness measure from the big five model was correlated with both dimensions of compliance in both treatments. There is limited evidence on the role of personality characteristics on behavior in experimental economics. Alaheto (2003) had found some correlation evidence between personality characteristics and white collar crime. Choo

et al. (2016) find evidence of a correlation between some big five personality traits and compliance on a sample of students, employees and self-employed. Curiously, the one personality trait they did not find evidence for was conscientiousness, for which we find strong evidence. In their meta-analysis of experiments on tax compliance, Alm and Malezieux (2021) state the evidence on the relationship between personality traits and deception is inconclusive. This question is certainly under-explored; it should be the focus of more research.

Our evidence on risk aversion and compliance points to the predicted negative direction, although we did estimate a surprising inconsistency in one estimated coefficient. In their meta-analysis of compliance, Alm and Malézieux (2021) find a very small impact of risk aversion on the decision to evade or not (extensive margin), but no impact on the extent to which participants evaded (intensive margin). Measurement error could be one reason why we find inconsistencies in the relationship between inferred risk aversion and compliance. There are many reasons why we might observe measurement error. First, the utility function assumed by the risk elicitation method could be different to that implied by participants' preferences. Second, the stakes involved in the decision may be too small for participants to care about, therefore not giving them enough of an incentive to precisely state their preferences. Third, participants may be fatigued at the end of the experiment, and as such be more prone to making genuine errors. To counter these considerations we would need to collect risk attitudes multiple times, and perhaps using a variety of methods. This was unfortunately beyond the scope of our study.

The fact that our experiment involves a real-effort task allows us to explore the inevitable heterogeneity in ability and how the resulting productivity differences map on tax evasion. We find that more productive individuals are also less compliant. However, windfall income does not lead to differences in compliance for differently productive individuals.

Our results should be interpreted with the caveat that our experiment was neutrally-framed: the language of the experiment avoided terms like "universal basic income" or "minimum guaranteed income". In reality, governments would frame such policies in language which could prompt reciprocity concerns on the part of citizens. Therefore, they might return a kind act with higher compliance. This psychological dimension of the implementation of a UBI program is important but perhaps beyond the scope of our paper. Likewise, and as we mentioned earlier, in a field context it is highly probable that guaranteed income from a government program could be assigned to a different mental account. Investigating these dimensions of behavior is certainly a worthwhile avenue for future research, probably best done in the field.

5 Conclusion

Universal basic income is once again at the forefront of public policy debates around the world. While the evidence around its effectiveness is rapidly increasing, questions remain about its consequences if implemented on the full population. One of them pertains to tax compliance, especially given that there is an increasing proportion of the labor force working as freelancers, or for whom third-party reporting is not complete. Should tax compliance and labor supply be adversely affected by large scale unconditional cash transfer programs such as UBI, this could dent the fiscal sustainability of such programs.

Our experiment makes a novel contribution to the behavioral literature on the effects of unconditional cash transfers by examining their effect on both labor supply as proxied by effort in a real-effort task, as well as tax compliance. Furthermore, we are able to examine these effects across the spectrum of labor productivity, unlike field studies which have focused on a restricted sub-set of the population. Furthermore, we make an additional contribution by examining these behaviors in Afghanistan, a country with incipient and fragile institutions, in which the state has had a limited history of public good provision, and very weak enforcement capabilities. We therefore extend the evidence on tax compliance and labor supply decisions on non-traditional samples.

We find windfalls have a negative effect on effort, particularly marked when windfalls were non-taxable (and therefore more salient). They also had a negative effect compliance that pertains to labor income, irrespective of whether windfalls are taxed or not. The estimated coefficients on windfall levels suggest a very similar effect on compliance across both treatments, which suggests we can rule out the possibility that having to declare both sources of income leads to changes in the profile of evasion. Interestingly, going from a medium sized to a large windfall results in an increase in average compliance as well as on the propensity to be fully compliant.

References

[1] Aaberge, R. & Colombino, U. (2014). Labor supply models. in: Cathal O'Donoghue (ed.), Handbook of Microsimulation Modelling, volume 127, pages 167-221, Emerald Publishing Ltd.

[2] Abeler, J., Falk, A., Goette, L. & Huffman, D. (2011). Reference points and effort provision. American Economic Review, 101(2), 470-92.

[3] Akee, R. K., Copeland, W. E., Keeler, G., Angold, A. & Costello, E. J. (2010). Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. American Economic Journal: Applied Economics, 2(1), 86-115.

[4] Akseer, T., Haidary, M. S., Miller, R., Sadat, S. M., Satkowski, C., Seese, H., ... & Yousufzai, F. A. (2018). Afghanistan In 2018: A Survey of the AFNhan People. *The Asian Foundation.*

[5] Akseer, T., Haidary, M. S., Miller, R., Sadat, S. M., Satkowski, C., Seese, H., ... & Yousufzai, F. A. (2017). Afghanistan In 2017: A Survey of the AFNhan People. *The Asian Foundation.*

[6] Alaheto, T. (2003). Economic crime: does personality matter? International Journal of Offender Therapy and Comparative Criminology, 47(3), 335-355.

[7] Allingham, M. G. & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. Journal of Public Economics, 1(3-4), 323-338.

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

[9] Alm, J. & Malezieux, A. (2021). 40 years of tax evasion games: a meta-analysis. Experimental Economics, 24, 699-750.

[10] Arkes, H. R. & Blumer, C. (1985). The psychology of sunk cost. Organizational Behavior and Human Decision Processes, 35 (1), 124-140.

[11] Ashenfelter, O. & Plant, M.W. (1990). Nonparametric estimates of the labor-supply effects of negative income tax programs. Journal of Labor Economics 8(1), S396-S415.

[12] Burtless, G. (1986). The work response to a guaranteed income: A survey of experimental evidence. In Conference Series Proceedings, No. 30, pp. 22-59. Federal Reserve Bank of Boston.

[13] Camerer, C., Babcock, L., Loewenstein, G. & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. The Quarterly Journal of Economics, 112(2), 407-441.

[14] Cesarini, D., Lindqvist, E., Notowidigdo, M. J. & Östling, R. (2017). The effect of wealth on individual and household labor supply: evidence from Swedish lotteries. American Economic Review, 107(12), 3917-46.

[15] Charness, G., Gneezy, U. & Halladay, B. (2016). Experimental methods: Pay one or pay all. Journal of Economic Behavior and Organization, 131, 141-150.

[16] Charness, G. & Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? In Handbook of Labor Economics, vol. 4, Part A, 229-330. Elsevier.

[17] Choo, C. L., Fonseca, M. A. & Myles, G. D. (2016). Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment. Journal of Economic Behavior and Organization, 124, 102-114.

[18] Chou, Y. K. (2002). Testing alternative models of labor supply: Evidence from taxi drivers in Singapore. The Singapore Economic Review, 47(1), 17-47.

[19] Collins, J., Murphy, D. & Plumlee, D. (1992). The taxpayer's dilemma: How hard to work and what to report. Advances in Taxation, 4, 31-53.

[20] Dai, Z., Galeotti, F. & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. Management Science, 64(3), 1081-1100.

[21] Davis, L., Joyce, B. & Roelofs, M. (2010). My money or yours: house money payment effects. Experimental Economics, 13 (2), 189-205.

[22] Doerrenberg, P. & Duncan, D. (2014). Experimental evidence on the relationship between tax evasion opportunities and labor supply. European Economic Review, 68, 48-70.

[23] Dwenger, N., Kleven, H., Rasul, I. & Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. American Economic Journal: Economic Policy, 8(3), 203-32.

[24] Dürsch, P., & Müller, J. (2017). Bidding for nothing? The pitfalls of overly neutral framing. Applied Economics Letters, 24(13), 932-935.

[25] Erard, B. & Feinstein, J. S. (1994). Honesty and evasion in the tax compliance game. The RAND Journal of Economics, 1-19.

[26] Fabre, A., Pallage, S. & Zimmermann, C. (2014). Universal Basic Income versus Unemployment Insurance. IZA Discussion Papers No. 8667.

[27] Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), 171-178.

[28] Frey, B. S. & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. The American Economic Review, 87(4), 746-755.

[29] Gill, D. & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. American Economic Review, 102(1), 469-503.

[30] Gneezy, U. & Potters, J. (1997). An experiment on risk taking and evaluation periods. The Quarterly Journal of Economics, 112(2), 631-645.

[31] Goette, L., Huffman, D. & Fehr, E. (2004). Loss aversion and labor supply. Journal of the European Economic Association, 2(2-3), 216-228.

[32] Guiso, L. & Paiella, M. (2008). Risk aversion, wealth, and background risk. Journal of the European Economic Association, 6(6), 1109-1150.

[33] Halla, M. (2012). Tax morale and compliance behavior: First evidence on a causal link. The BE Journal of Economic Analysis & Policy, 12(1). https://doi.org/10.1515/1935-1682.3165

[34] Hamal, K. B. & Anderson, J. R. (1982). A note on decreasing absolute risk aversion among farmers in Nepal. Australian Journal of Agricultural Economics, 26(3), 220-225.

[35] Henrich, J., Heine, S. J. & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. Behavioral and Brain Sciences, 33(2-3), 111.

[36] Hum, D. & Simpson, W. (1993). Economic response to a guaranteed annual income: Experience from Canada and the United States. Journal of Labor Economics, 11(1, Part 2), S263-S296.

[37] Imbens, G. W., Rubin, D. B. & Sacerdote, B. I. (2001). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. American Economic Review, 91(4), 778-794.

[38] Juodis, A., Karavias, Y. & Sarafidis, V. (2021). A homogeneous approach to testing for Granger non-causality in heterogeneous panels. Empirical Economics, 60(1), 93-112.

[39] Jones, D. & Marinescu, I. (2018). The labor market impacts of universal and permanent cash transfers: Evidence from the Alaska Permanent Fund (No. w24312). National Bureau of Economic Research.

[40] Kessler, J.B. & Norton, M.I. (2016). Tax aversion in labor supply. Journal of Economic Behavior and Organization, 124, 15-28.

[41] Kőszegi, B. & Rabin, M. (2006). A model of reference-dependent preferences. The Quarterly Journal of Economics, 121(4), 1133-1165.

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

[43] McCaffery, E. J. & Baron, J. (2006). Thinking about tax. Psychology, Public Policy, and Law, 12(1), 106.

[44] Marinescu, I. (2018). No strings attached: The behavioral effects of US unconditional cash transfer programs (No. w24337). National Bureau of Economic Research.

[45] Mascagni, G. (2018). From the lab to the field: A review of tax experiments. Journal of Economic Surveys, 32(2), 273-301.

[46] Olsen, J., Kogler, C., Brandt, M. J., Dezső, L., & Kirchler, E. (2019). Are consumption taxes really disliked more than equivalent costs? Inconclusive results in the USA and no effect in the UK. Journal of Economic Psychology, 75, 102145.

[47] Ortiz, I., Behrendt, C., Acua-Ulate, A. & Anh, N. Q. (2018). Universal Basic Income proposals in light of ILO standards: Key issues and global costing. Available at SSRN 3208737.

[48] van Parijs, P. & Vanderborght, Y. (2017). Basic income. A radical proposal for a free society and a sane economy. Cambridge, MA: Harvard University Press.

[49] Pencavel, J. H. (1979). A note on income tax evasion, labor supply, and nonlinear tax schedules. Journal of Public Economics, 12(1), 115-124.

[50] Picchio, M., Suetens, S. & van Ours, J.C. (2018). labor supply effects of winning a lottery. Economic Journal 128(611), 1700-1729.

[51] Robins, P. K. (1985). A comparison of the labor supply findings from the four negative income tax experiments. Journal of Human Resources, 20(4), 567-582.

[52] Sapsford, D. & Tzannatos, Z. (1993). The economics of the labor market. Macmillan International Higher Education.

[53] Stern, A. (2016). Raising the floor: How a universal basic income can renew our economy and rebuild the American dream. Public Affairs.

[54] Sussman, A. B. & Olivola, C. Y. (2011). Axe the tax: Taxes are disliked more than equivalent costs. Journal of Marketing Research, 48(SPL), S91-S101.

[55] Thaler, R. & Johnson, E. (1990). Gambling with the house money and trying to break-even: The effects of prior outcomes on risky choice. Management Science, 36(6), 643-660.

[56] Torgler, B. (2007). Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis, Cheltenham, UK: Edward Elgar.

[57] Torgler, B. & Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. Journal of Economic Psychology, 30(2), 228-245.

[58] US Federal Reserve (2019). Report on the Economic Well-Being of US Households in 2018, May 2019. Board of Governors of the Federal Reserve System, Washington, DC.

[59] Woods, S. A. & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. European Journal of Personality: Published for the European Association of Personality Psychology, 19(5), 373-390.

A Econometric Analysis

B Distribution of Compliance

We report the distribution of compliance (labor and windfall) for the TW and NTW treatments.

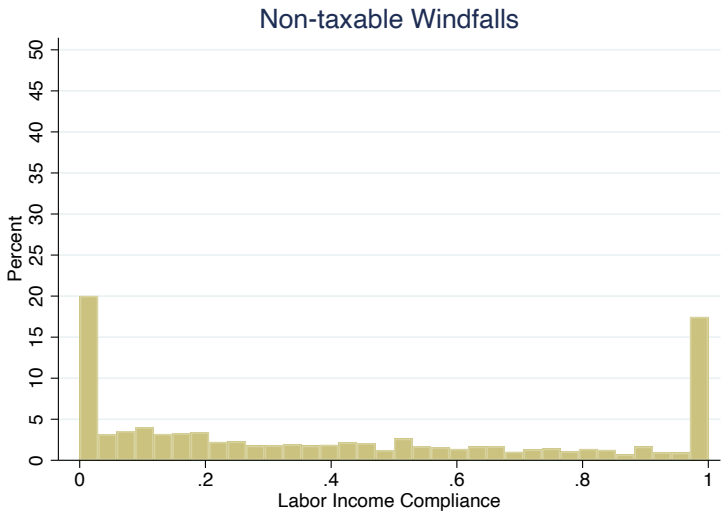

Figure 1: Labor Compliance, NTW treatment

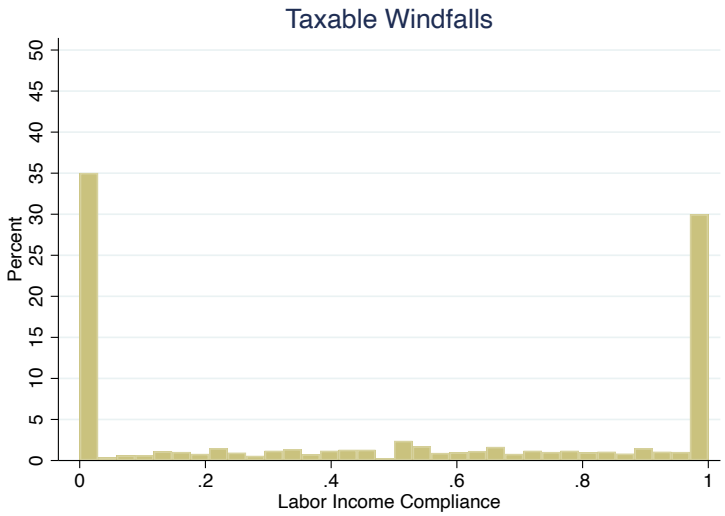

Figure 2: Labor Compliance, TW treatment

	TaxableWindfall	NonTaxableWindfall	p-value
Male	0.61	0.54	0.172
Age	22.23	25.00	<0.0001
	(2.24)	(3.10)	
Risk	3.00	2.46	<0.0001
	(1.03)	(0.89)	
Extraversion	4.46	4.68	0.397
	(2.71)	(2.78)	
Agreeableness	6.13	6.78	0.007
	(2.63)	(2.40)	
Emotional stability	4.43	5.00	0.037
	(2.85)	(2.87)	
Conscientiousness	3.78	4.12	0.192
	(2.66)	(2.73)	
Openness	5.33	5.80	0.076
	(2.70)	(2.67)	
Attitude toward the government	5.32	5.43	0.725
	(3.14)	(3.16)	
Mistrust in government	4.22	4.20	0.933
	(2.82)	(2.93)	
Knows evasion	5.80	6.18	0.196
	(3.01)	(3.17)	
Tax morale	7.20	6.83	0.158
	(2.60)	(2.71)	
N	215	215	

Note: p-values are based on two-sample t-test for difference in means in the case of non-dichotomous variables; Fisher's exact test in the case of dichotomous variables.

Table 5: Sample characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Extraversion	1.00											
(2) Agreeableness	-0.04	1.00										
(3) Emotional stability	0.24	0.05	1.00									
(4) Conscientiousness	0.13	0.05	-0.13	1.00								
(5) Openness	-0.09	0.22	0.01	-0.01	1.00							
(6) Attitude toward the gov't	0.14	0.03	0.00	0.03	0.12	1.00						
(7) Tax morale	0.14	0.14	-0.02	0.07	0.03	0.29	1.00					
(8) Lack of trust in gov't	0.06	0.00	0.13	0.05	-0.12	-0.00	-0.06	1.00				
(9) Knows evasion	-0.08	0.24	-0.05	0.04	-0.04	0.20	0.27	-0.01	1.00			
(10) Risk Attitude	-0.03	-0.05	0.04	0.01	0.04	0.06	-0.00	-0.04	-0.06	1.00		
(11) Age	-0.01	0.13	-0.02	0.03	0.07	0.00	0.04	-0.07	0.05	-0.19	1.00	
(12) Male	0.02	-0.04	-0.04	-0.01	-0.05	0.06	0.01	0.11	-0.07	0.01	-0.01	1.00

Table 6: Sample correlations of regressors

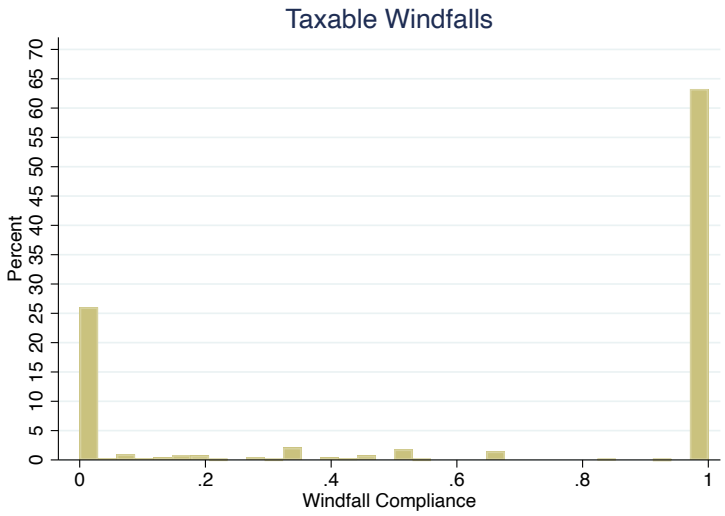

Figure 3: Windfall Compliance, TW treatment

B.1 Testing for the Simultaneity of Effort and Compliance

In this sub-section we provide evidence for the claim that participants in our experiment made their effort decision independently of their compliance decision and decided on compliance condition on their effort decision. To do this, we conducted a 2SLS analysis where in the first stage we instrumented labor income compliance using tax morale. In the second stage, we regressed effort on estimated compliance and the two windfall levels. If the decision to comply is done simultaneously with effort, we should observe a significant coefficient on the compliance variable in the second stage regression. In both cases, the coefficient on labor income is non-significant in either specification, supporting our conjecture.

In addition to this test, we performed a Granger causality test of effort and labor income compliance based on Juodis, Karavias and Sarafidis (2021). We had to remove from this analysis any participant who had near-zero variance in compliance, as the test requires positive variance in both variables. This meant our working sample was 121 individuals. We failed to reject the null of no Granger-causality: HPJ Wald test $.95015782, p = 0.3297$.

First stage estimates		
DV: Labor compliance	Taxable Windfall	Non-taxable Windfall
Tax morale	0.019**	0.022***
	(0.008)	(0.007)
Windfall =15	-0.016	0.011
	(0.011)	(0.011)
Windfall =30	-0.007	-0.005
	(0.011)	(0.012)
Constant	0.353***	0.270***
	(0.063)	(0.054)
F test of excluded instruments	$F(1, 214) = 5.14$	$F(1, 214) = 9.16$
p-value	0.024	0.003
Second stage estimates		
DV: Effort	Taxable Windfall	Non-taxable Windfall
labor income compliance	15.867	-1.167
	(11.182)	(6.097)
Wf15	-0.030	-0.513***
	(0.237)	(0.195)
Wf30	-1.527***	-0.701***
	(0.275)	(0.191)
Constant	6.669	16.472***
	(5.499)	(2.612)
Kleibergen-Paap LM statistic	4.744	9.380
Kleibergen-Paap LM p-value	0.029	0.002
Cragg-Donald Wald F statistic	34.947	69.157
Weak ID test critical values: 10% maximal IV size	16.38	16.38
N	2,580	2,580

Note: Kleibergen-Paap LM statistic is an underidentification test; the Cragg-Donald Wald F test is a weak identification test. We report the statistic and the critical value for 10% maximal IV bias.

Table 7: Test of sequentiality

	Taxable Windfall		Non-taxable Windfall	
Wf15	-0.041	0.006	-0.017	-0.039
	(0.061)	(0.112)	(0.033)	(0.058)
Wf30	-0.040	-0.072	-0.033	-0.075
	(0.060)	(0.111)	(0.032)	(0.058)
Ability		-0.075***		-0.010
		(0.016)		(0.006)
Tercile 2	-0.534***		-0.067	
	(0.184)		(0.073)	
Tercile 3	-0.796***		-0.167**	
	(0.192)		(0.075)	
Wf15 × Tercile 2	0.052		-0.034	
	(0.084)		(0.044)	
Wf15 × Tercile 3	-0.108		-0.008	
	(0.087)		(0.045)	
Wf30 × Tercile 2	0.106		-0.012	
	(0.083)		(0.044)	
Wf30 × Tercile 3	-0.018		-0.022	
	(0.086)		(0.045)	
Wf15 × Ability		-0.004		0.000
		(0.007)		(0.003)
Wf30 × Ability		0.004		0.003
		(0.007)		(0.003)
Extraversion	-0.028	-0.033	0.003	0.002
	(0.030)	(0.029)	(0.011)	(0.011)
Agreeableness	0.005	0.010	-0.003	-0.004
	(0.029)	(0.028)	(0.014)	(0.014)
Emot. Stability	0.030	0.039	-0.027***	-0.026**
	(0.028)	(0.028)	(0.010)	(0.010)
Conscientiousness	0.077***	0.079***	0.020	0.020
	(0.029)	(0.029)	(0.011)	(0.011)
Openness	-0.040	-0.043	0.010	0.011
	(0.027)	(0.026)	(0.012)	(0.012)
Att. Towards Gov't	-0.019	-0.019	-0.017	-0.017
	(0.026)	(0.026)	(0.010)	(0.010)
Tax Morale	0.078**	0.086***	0.027**	0.027**
	(0.032)	(0.032)	(0.012)	(0.012)
Mistrust in Gov't	-0.043	-0.040	-0.007	-0.007
	(0.027)	(0.027)	(0.010)	(0.010)
Knows Evade	-0.028	-0.032	-0.007	0.003
	(0.026)	(0.025)	(0.010)	(0.010)
Risk tolerance	-0.144**	-0.124	0.027	0.023
	(0.072)	(0.071)	(0.032)	(0.032)
Age	0.024	0.023	0.012	0.012
	(0.033)	(0.033)	(0.009)	(0.009)
Male	-0.023	-0.007	0.067	0.069
	(0.150)	(0.148)	(0.058)	(0.058)
Period	-0.007	-0.007	0.015***	0.015***
	(0.004)	(0.004)	(0.002)	(0.002)
Constant	0.669	1.180	-0.067	0.029
	(0.846)	(0.846)	(0.285)	(0.292)
N	2,580		2,580	
LL	-1780.6	-1780.3	-1599.1	-1600.4
χ^2	57.67	59.04	80.22	77.41

Note Standard errors clustered at the individual level. Wf15 corresponds to Low condition (15 tokens) and Wf30 to High condition treatment (30 tokens). ***, **: $p < 0.01, p < 0.05$ respectively.

Table 8: Impact of windfalls on labor income compliance for different levels of ability

C Instructions

Welcome to today's experiment at Rana University. You will be taking part in an experiment on decision-making. Starting from now, you are kindly asked to refrain from communicating with other participants in this session. You will be paid for your time and effort; your cash earnings depend solely on your decisions. You have been provided with a hardcopy of the instructions, you may refer to these at any point during the experiment or alternatively you can raise your hands and the experimenter will answer your questions individually without disturbing others. It is important to understand the rules of the experiment before you proceed so pay close attention to the following instructions please. In this experiment, your earnings are calculated in tokens and 1 token is equal to 30 Afs. At the end of the experiment, the average of 3 randomly rounds are chosen for payment purposes. Your token earnings will be converted, rounded up to ten into AFNhanis (Afs) and will be paid privately in cash. There are two parts in this experiment.

Part A:

Stage 1:

The amount of variable tokens is publicly announced to the participants. After hearing the variable token, record these tokens in the box provided in the z-tree screen. In order to save and proceed to the following stage each time, you must click confirm.

Stage 2:

You are asked to solve 48 slider puzzles in order to earn more tokens. You are paid at a piece rate of 1 token per correctly positioned slider puzzle. Each slider is positioned at either end of the line (0 or 100) and your tasked to drag and place this at 50 using a mouse. Only correctly positioned sliders will earn you tokens. You have also been given the option to opt-out should you wish. You can do so by clicking the SKIP red-button to opt-out of this stage at any point during the task without losing any of your earnings. You have 120 seconds to complete this task.

Stage 3:

All your earnings are taxable and you will be presented with a tax form to report your gross income for each round. There is a tax rate of 35% on your earnings, for example for 10 tokens, 3.5 tokens taxes will apply. After you submit your tax returns, the correct amount is calculated by the computer and is deducted from your total earnings. There is also a 5% chance of being selected for tax audit:

- If you are audited and if you have reported your earnings accurately, then no further action will take place. Your final earnings will stay the same as before being audited.
- If you are audited and if you have underreported your earnings, in addition to the correct amount of the tax, you will pay a fine equal to 25% of the unpaid taxes. The audit probability is independent in each round and your previous filing behaviour doesn?t change the audit probability for the next round.

Stage 4:

The final screen will present the following information for that round:

- Your variable tokens
- Your tokens earned from the slider puzzles
- Your declared variable tokens
- Your declared slider task tokens
- Paid amount in taxes
- Net earnings (after tax deductions)
- Whether you were audited or not:
- Your final payoff for that round in tokens

Rounds repeat themselves after stage 4, i.e. after stage 4, it goes back to stage 1 and you are required to scratch a new card for the variable tokens. The average of 3 random rounds from this is chosen for payment.

Part B:

Stage 5:

You will be asked to make an investment decision in the following manner: you will have 4 tokens and you are required to invest this in an account that pays 2.5 tokens for every token invested with a probability 1/2; however, with probability 1/2 it pays back nothing. To determine the outcome, the computer will draw a number between 1 and 10; if numbers 1-5 are picked, then you lose the amount you allocated; if numbers 6-10 are picked, then you win 2.5 times the amount you allocated. The unallocated amount from the 4 original tokens and the outcome of this lottery will be added to your final tokens in the experiment.

Stage 6:

Finally, you will be presented with a set of questions asking you to state on a scale of 1 to 9 how much the statement applies to you; 1 being very confident and 9 being least confident. With this round, the experiment comes to an end.

1. It is acceptable to not declare some of your income to the tax authority (e.g. income from a second job).
2. There is no point cheating as you will always get caught.
3. I am aware of ways to cheat the system.
4. I am confident I pay all the tax that I owe.

Stage 7:

You will be notified of your final earnings in both tokens and AFNhanis (Afs). You are required to confirm this by clicking the button at the bottom of the page. Once this is done, please notify the experimenter by raising your hand and wait to be attended to for payment.

D Sample Screenshots – NTW treatment

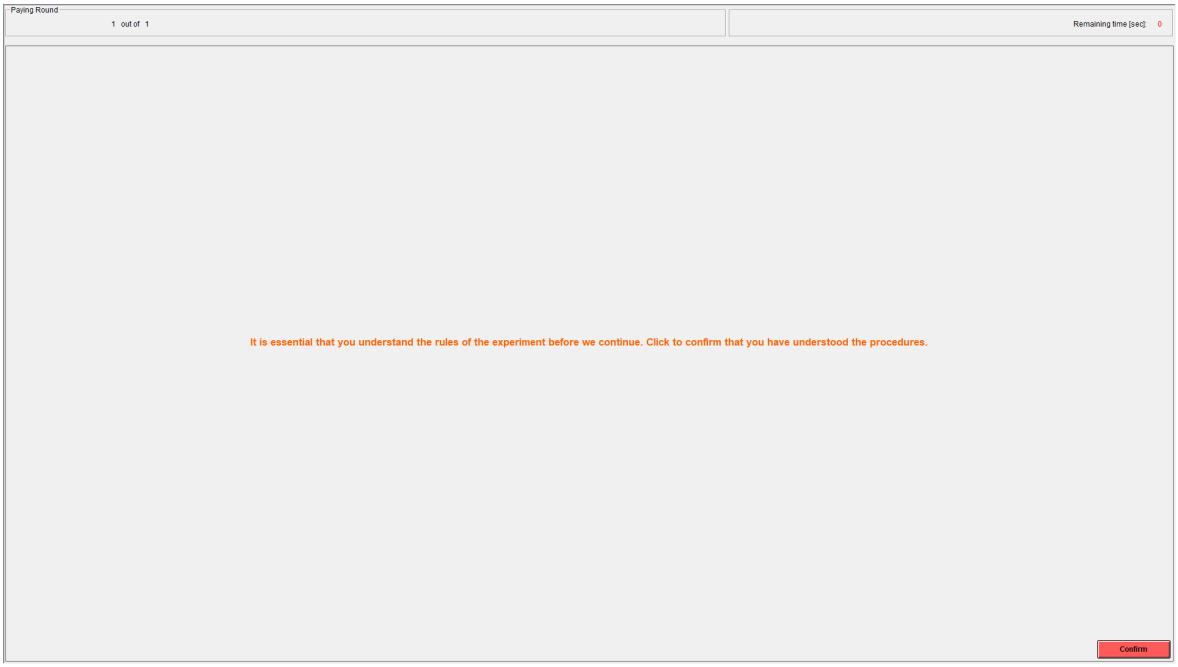

Figure 4: Comprehension acknowledgment screen. Subjects clicked button after asking any clarification questions to experimenter.

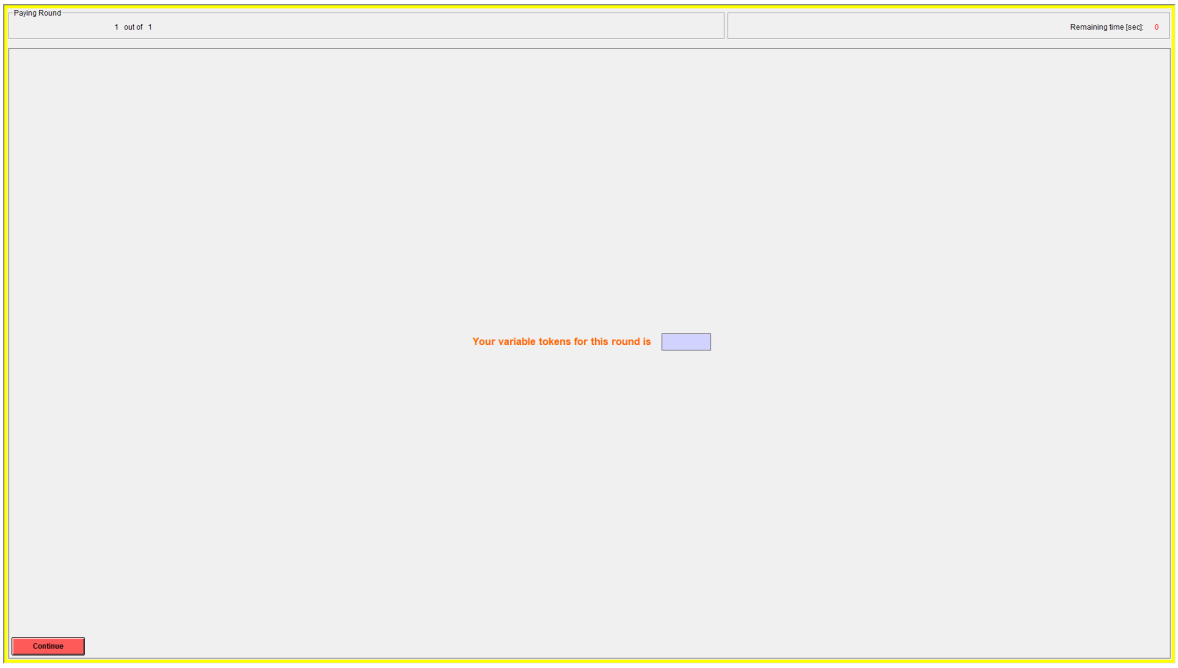

Figure 5: Variable tokens (windfall) screen. Subjects had to enter the amount announced by the experimenter. Subjects could not enter a different amount on the box.

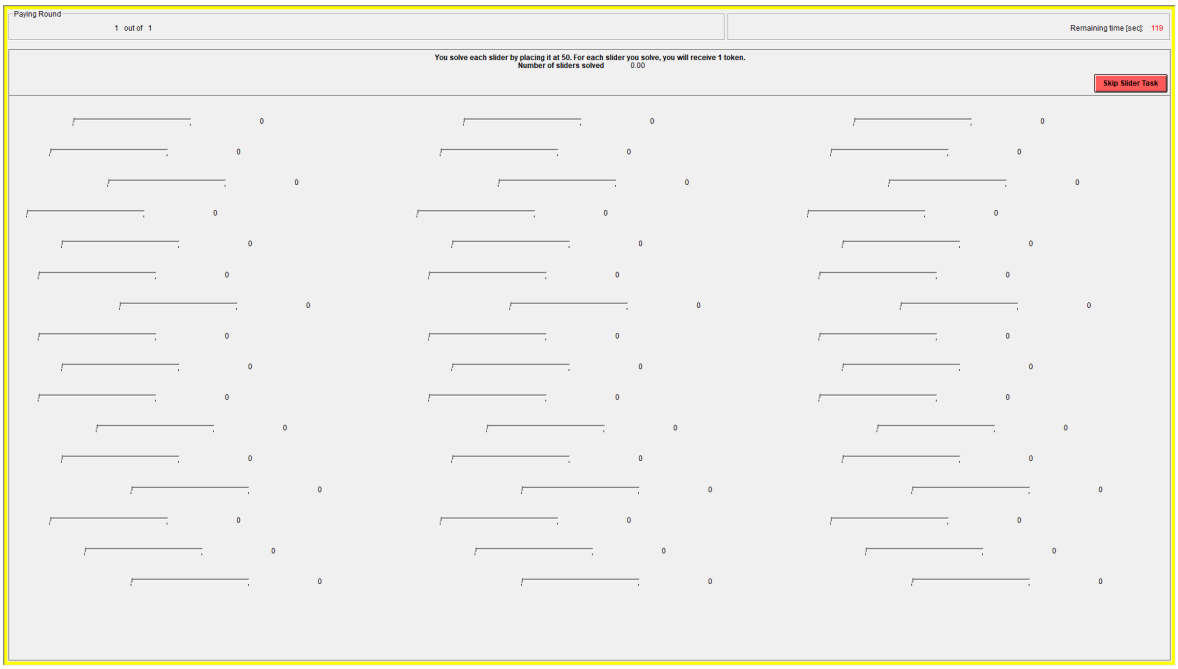

Figure 6: Slider task screen. Subjects solve as many sliders as they wished within a time limit for a piece rate of 1 token per slider.

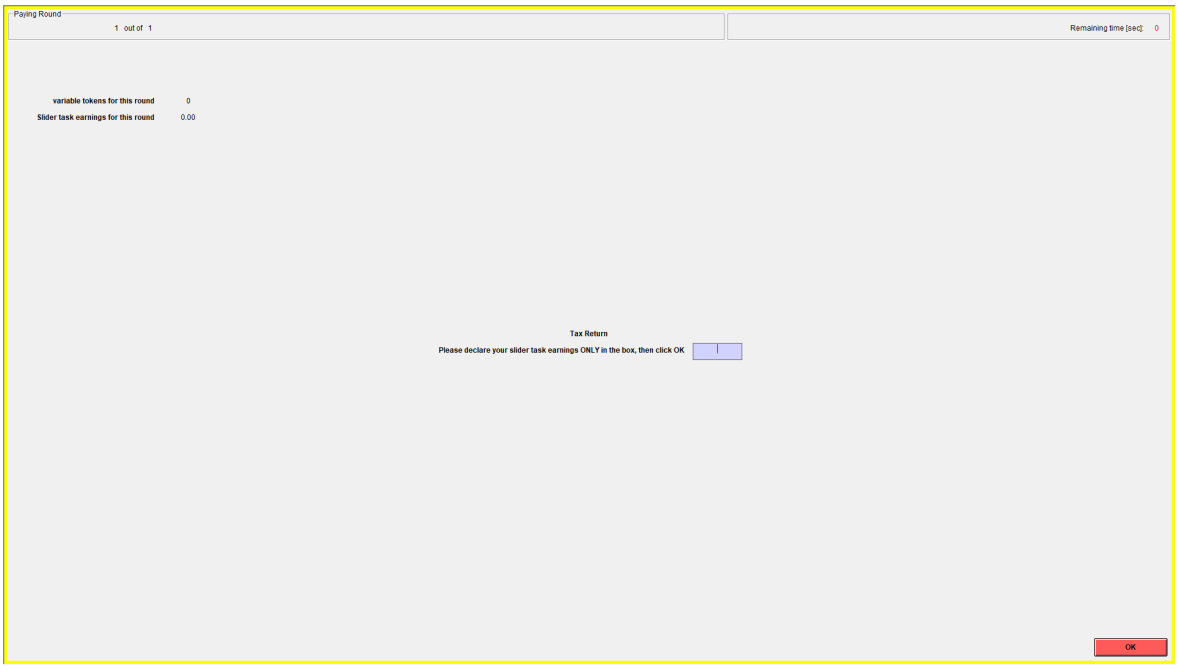

Figure 7: Slider income declaration screen, NTW treatment.

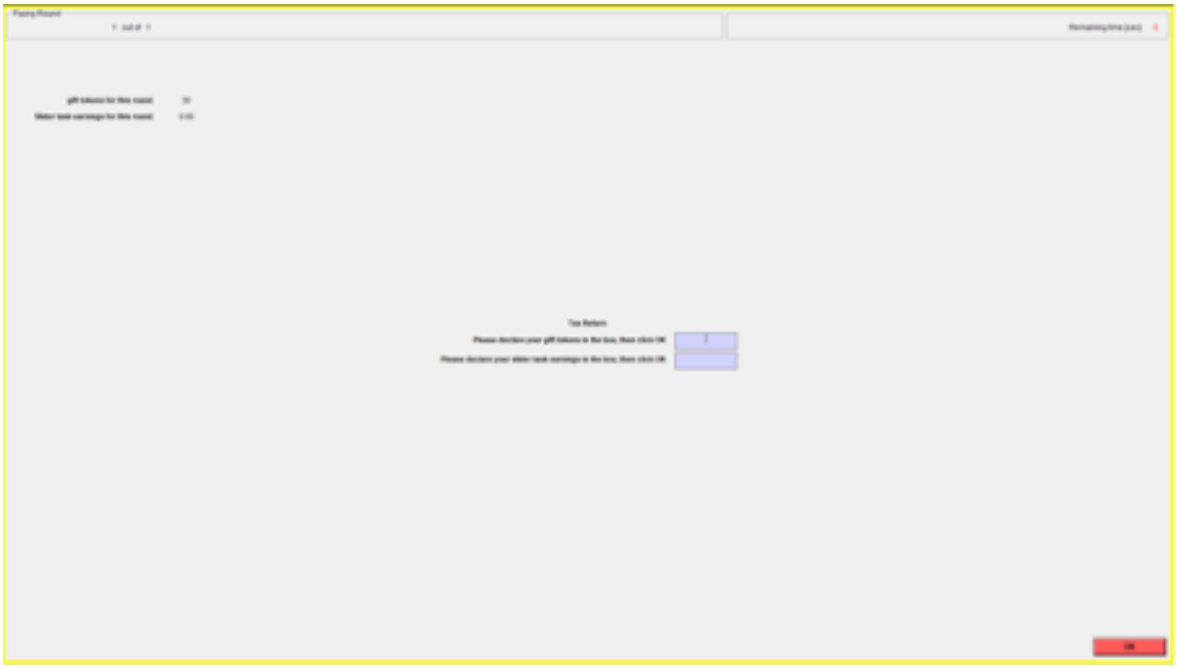

Figure 8: Slider income declaration screen, TW treatment.

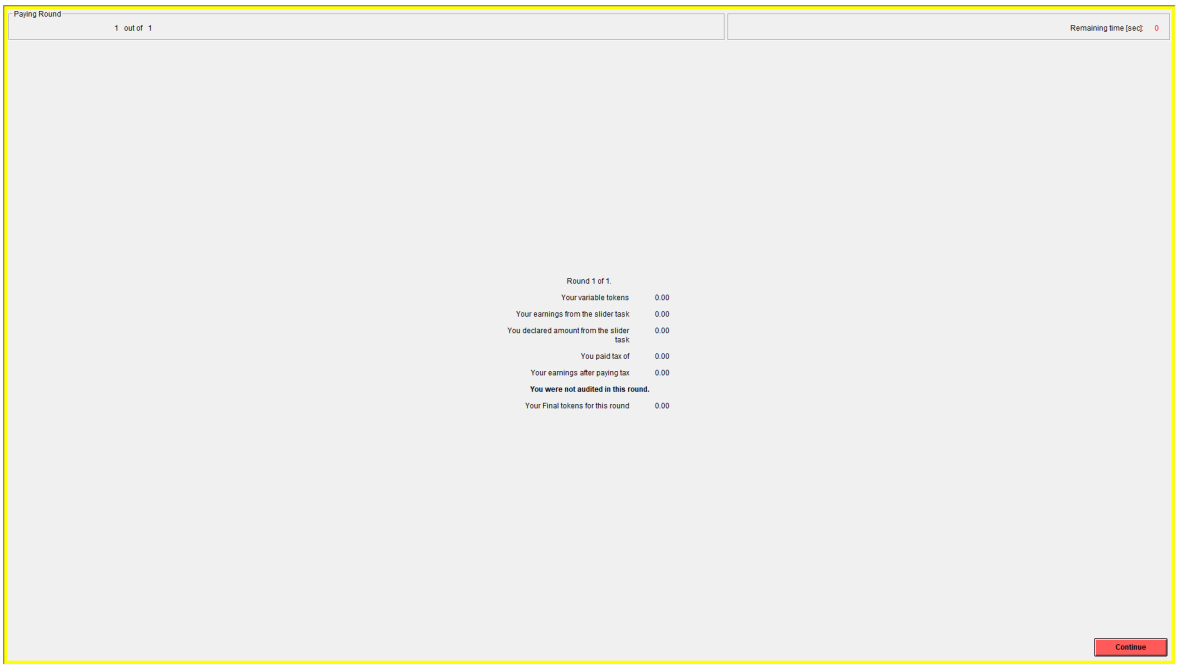

Figure 9: End of round feedback screen.

Figure 10: Risk attitude elicitation screen.