Journal of Economic Psychology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

Are consumption taxes really disliked more than equivalent costs? Inconclusive results in the USA and no effect in the UK

Jerome Olsen[a,*], Christoph Kogler[b], Mark J. Brandt[b], Linda Dezső[c], Erich Kirchler[a]

[a] University of Vienna, Faculty of Psychology, Department of Applied Psychology: Work, Education and Economy, Universitaetsstrasse 7, 1010 Vienna, Austria

[b] Tilburg University, School of Social and Behavioral Sciences, Department of Social Psychology, Warandelaan 2, 5037AB Tilburg, the Netherlands

[c] University of Vienna, Vienna Center for Experimental Economics, Oskar-Morgenstern-Platz, 1, 1090 Vienna, Austria

ARTICLE INFO

Keywords:
Tax aversion
Replication
Tax behavior
VAT
Sales tax
Financial choice

ABSTRACT

In two experiments on hypothetical purchase decisions, Sussman and Olivola (2011) found that US citizens prefer avoiding tax-related costs over avoiding tax-unrelated monetary costs of the same size. The original Experiment 1 and 2 tests of this Tax Aversion indicated that people are willing to wait longer to receive a discount when it refers to taxes (e.g., "axe-the-tax discount") than when it is just a regular discount (e.g., "customer rewards"). We conducted high-powered close replications of both original studies, Experiment 1 ($N = 590$) and Experiment 2 ($N = 650$), which reveal either no effect (Experiment 1: $r = 0.02$, 95% CI [−0.06, 0.10]) or a small effect (Experimental 2: $r = 0.09$, 95% CI [0.01, 0.16]) in the USA. We also replicated both experimental procedures in the UK to test whether the effect generalized to a value added tax system. Neither Experiment 1 ($N = 595$; $r = 0.01$, 95% CI [−0.07, 0.09]) nor Experiment 2 ($N = 673$; $r = 0.03$, 95% CI [−0.04, 0.11]) revealed an effect in the UK. Tax Aversion in hypothetical consumption decisions seems to be a smaller phenomenon than originally proposed and does not generalize to a value added tax system.

1. Introduction

Calling a financial charge on citizens by what it is – a tax – or using a mild alternative – e.g., community charge – can lead to different perceptions by the public. One historical account is the Community Charge in the UK during the third and last premiership of Margaret Thatcher. Prior to 1989, local governments collected so-called Domestic Rates, a tax based on the property value of citizens' houses. Thatcher's 1989 reform introduced a flat tax on property owners and called it a Community Charge. The opposition and the media called it a Poll Tax instead of Community Charge, which some have credited with, at least in part, the Community Charge's quick demise (for detailed discussions, see Bramley, Grand, & Low, 1989; Smith, 1991; Winetrobe, 1992).

Along these lines, researchers have identified how framing information can affect perceptions and attitudes. For instance, experimental research revealed a higher willingness to pay for an equal-sized environmental charge when it was labeled as Carbon *Offset* than when labeled as Carbon *Tax* (Hardisty, Johnson, & Weber, 2010), suggesting an aversion to taxes. Mccaffery and Baron (2006) identify such inconsistent evaluations of tax policies as specific instances of a more general isolation or focusing effect. That is, citizens' decisions on complex matters are often based on the most salient or obvious aspect of a decision problem. This partly explains why logically relevant information is often ignored and evaluations are influenced by easily processed labels or metrics echoed in public discourse.

* Corresponding author.
E-mail address: jerome.olsen@univie.ac.at (J. Olsen).

https://doi.org/10.1016/j.joep.2019.02.001
Received 1 June 2018; Received in revised form 1 February 2019; Accepted 2 February 2019

1.1. The original axe the tax study

Sussman and Olivola (2011) demonstrate a similar phenomenon they label Tax Aversion, which they define as a preference for avoiding tax-related costs over monetary costs of equal size (or even higher) that are not related to taxes. Their study consists of a series of five independent experiments, investigating Tax Aversion in different contexts. In Experiments 1 and 2 they manipulated the framing of a purchase discount as either related to tax (e.g., "axe-the-tax") or as a regular, tax-unrelated discount (e.g., "customer rewards") between subjects. In Experiment 1 participants had to indicate whether they would be willing to accept a longer car drive to receive the discount, while in Experiment 2 individuals had to state their maximum willingness to wait in minutes to receive the discount. Results of both experiments showed that people are willing to wait (drive or stand in line) longer for a tax-related versus a tax-unrelated discount. Experiment 2 additionally revealed that consumers who were offered a tax-related discount tended to ask for smaller discounts for standing in line for 15 minutes than those offered a tax-unrelated discount. The remaining three experiments showed that individuals prefer tax-exempt bonds over equally profitable bonds that are subject to tax (Experiment 3), that Tax Aversion is higher among individuals who identify with antitax parties (Experiment 4), and that this latter effect is mitigated if antitax sympathizers are instructed to reflect on positive uses of taxes (Experiment 5).

Results of these five experiments demonstrate that citizens' tax attitudes influence decisions beyond the immediate context of tax behavior and extend to different decisions in daily life. Accordingly, the authors discuss their results not only in terms of implications for tax policy but also for marketers. One methodological strength of the reported experiments is that they may be less prone to social desirability as compared to studies that explicitly ask participants about their tax attitudes.

To evaluate the original study's impact, we ran a bibliometric search in three different databases (Google Scholar, Web of Knowledge, and Scopus) on April 18th 2018 and found 73 unique citations since the paper's publication in 2011. More than half of the citations have a clear focus on tax-related issues and the publication is most often cited to refer to citizens' general aversion to taxes. Given the attention the publication has received so far, we believe a close replication attempt of the phenomenon of Tax Aversion is important and thus conducted independent replications of Experiments 1 and 2 which capture the focal effects of the paper. In doing so, we followed the guidelines of the Replication Recipe proposed by Brandt et al. (2014).

1.2. Axe the tax mechanisms

The mechanism behind the observed preference for a tax-related over a tax-unrelated discount in Experiments 1 and 2 could be twofold. First, a general aversion to taxes could drive the effect, where just thinking of saving on taxes is sufficient. Second, in a sales tax system where taxes are added only at the check-out, consumption taxes constitute a salient out-of-pocket loss which could drive the effect (Chetty, Looney, & Kroft, 2009). That is, tax-related discounts imply that customers can actually avoid paying consumer tax (this is not the case) and so consumers are likely to compare the price plus sales tax (regular price) against the price without any additional tax (tax-related discount). We suspect that both explanations apply and that the latter might be one driving factor of a general aversion to taxes. This view is also in line with the fact that Sussman and Olivola (2011) find Tax Aversion in investment decisions – a different domain – in Experiment 3, where sales tax does not apply.

The idea that a focus on the regular price (including sales tax) compared to a price without any additional tax could drive this observed effect of Tax Aversion raises the question of whether customers in a value added tax system, where the tax is displayed as part of the price and should therefore be less salient (Bird, 2010), show the same preference for tax-related over tax-unrelated discount offers. Here individuals in both conditions are likely to compare the options based on the final sales price instead of either thinking of saving tax or receiving a discount on the pretax price. To obtain empirical evidence of whether the observed effect is specific to sales tax or rather a consequence of general Tax Aversion, and to expand the research question to consumption taxes in general, we extended the replication efforts to a country that uses value added tax, namely the UK.[1] We chose the UK because the materials required only small modifications (i.e., small language and cultural differences), and a replication attempt was therefore fairly easy to implement.

In summary, we conducted close replications of Experiment 1 and 2 of Sussman and Olivola (2011) in the USA to test whether Tax Aversion, operationalized through purchase decision scenarios, is a stable phenomenon. Additionally, we conducted both experiments in the UK to answer the question whether Tax Aversion is present in a value added tax system.

This manuscript was submitted as a registered report and therefore peer-reviewed prior to data collection (see https://osf.io/7pruq/ for the stage 1 submission). Further preregistrations prior to data collection were filed for decisions made after receiving peer-review comments (see https://osf.io/g2jbr/ for pre-tests and https://osf.io/nqg3e/ for additional exploratory analyses). The main OSF project website links to all these registrations and contains data, R code, and materials (https://osf.io/q8g7f/).

2. Experiment 1

2.1. Method

See Table S1 of the supplementary material for the 36-question guide to the Replication Recipe (Brandt et al., 2014) that

[1] In footnote 4 of the original study, Sussman and Olivola (2011) refer to a replication of their Experiment 4 carried out on a UK sample. Here they find Tax Aversion only among more right-leaning individuals.

addresses details about the nature of the original effect and differences between the original and replication study.

2.1.1. Participants

2.1.1.1. Original study. A total of 238 participants were recruited for the original study using three different strategies (n = 131 Amazon's Mechanical Turk (MTurk), n = 65 passersby in a US shopping mall, and n = 42 Princeton University undergraduates). After excluding 47 individuals (non-US residents, already participated in same study, or visibly challenged), the final sample size was N = 191 with 62% female participants and a mean age of M = 29.9 years (SD = 11.9).

In the original study, those recruited through MTurk and the shopping mall received a fixed payment for completion, while students received either course credits or payment for their participation. Respondents from the shopping mall and on campus filled in a survey package that also contained unrelated questionnaires. In the analysis these samples were merged as they did not differ from each other.

2.1.1.2. Power analysis. The focal effect was the proportion of respondents preferring to travel 30 minutes to receive a price discount on a new television compared to paying a regular price, where the discount was either an 8% tax-related discount or a 9% tax-unrelated discount (between-subject). In the original experiment, participants in the tax-related discount condition were more likely to accept the additional travel time for a discount than participants in the tax-unrelated discount condition (76% vs. 59%), $\chi^2(1, N = 191) = 5.83$, $p = .016$, $r_\Phi = 0.18$, 95% CI [0.04, 0.31].

Running a power analysis for differences between these two independent proportions using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) yielded a sample size of N = 390 for $r_\Phi = 0.18$ with $\alpha = 0.05$ and $1 - \beta$ (power) = 0.95. This assumes that the effect estimate in the original paper is an accurate estimate of the true underlying effect size. Given the general risk of an overestimated effect, our power could have then turned out to be too low.

To address this, two procedures have been proposed in the literature that adjust targeted effect sizes downward: safeguard power (Perugini, Gallucci, & Costantini, 2014) and the 2.5 × N rule (Simonsohn, 2015). In the first case, the original effect's lower-bound of the 60% confidence interval is used as a targeted effect size. The original effect size was $r_\Phi = 0.18$ with 60% CI [0.12, 0.24]. Using the lower-bound of $r_\Phi = 0.12$ would require N = 896 for 95% power. In the second case, it is suggested to multiply the original sample size by a factor of 2.5, which would yield N = 573. Given our available resources, we set the targeted sample size at N = 600 for a replication of Experiment 1 for both the US and UK sample, respectively. This provided 95% power for effects as small as $r_\Phi = 0.15$ and still 80% power for effects as small as $r_\Phi = 0.11$ in each replication sample.

2.1.1.3. Replication study samples and procedure. Participants were recruited using Prolific Academic which allowed sampling individuals in the USA and the UK. We successfully recruited N = 596 US and N = 600 UK participants. The only inclusion criterion was that individuals had to indicate being US/UK residents. After exclusions, the final samples were N = 590 for the USA and N = 595 for the UK. Women made up 51% and 67% of the sample in the USA and UK, respectively. Mean age was M = 34.4 years (SD = 11.7) in the USA and M = 37.6 years (SD = 12.3) in the UK. See Fig. S1 of the supplementary material for key demographics by sample and condition.

Data collection took place in late November 2018. Median participation time was slightly below three minutes in both samples. Individuals were paid a remuneration of $1.00 for completion.

2.1.2. Materials

2.1.2.1. Materials for US sample. All materials were available from the original publication. Furthermore, we contacted the original authors regarding further unreported though important procedural details to the study design. We were informed that this was not the case.

The study comprised a two-group between-subject design. Both groups read a scenario about a purchase decision for a new television. Participants had to decide between paying the regular price or accepting a higher time investment to receive a discount. The between-subject factor varied the type of discount (8% tax-related discount vs. 9% tax-unrelated discount). Participants read the following scenario, followed by making a decision for one of two stores that either offered the television at the regular price or with one of the two discounts.

> "You want to buy a new television and have a particular model in mind. Calling around, you find that only two stores, Bob's Electronics and Tom's Electronics, carry that model. Bob's Electronics is located very close, about a 5-minute drive, but offers no discounts on the television set. Tom's Electronics is located farther away, about a 30-minute drive, but offers the television set [tax-free, which is equivalent to an 8% discount/with a 9% discount]. Where do you go to make your purchase?"
>
> ☐ Bob's Electronics ☐ Tom's Electronics

In a pre-test we asked N = 51 US individuals (sampled from Prolific Academic) how long they would be willing to drive to receive a 9% discount. They answered binary yes/no questions for 5-minute time intervals increasing from 5 to 100 minutes. We pre-registered that we would use the pre-tested median driving time in the final scenario, which was 30 minutes as in the original study.

In the original study, participants were asked about their gender, age, and personal income. To further explore potential moderators of Tax Aversion, in the replication study, participants answered five question blocks after choosing between the two stores.

In the first block, participants were first given one item asking about their political ideology with a 7-point answering scale ranging from *extremely liberal* to *extremely conservative*, followed by another item asking about the party they most strongly identify

with. For the indicated party, they had to state the strength of their identification. Additionally, participants were asked about their satisfaction with their current government.

In the second block, participants were presented with four statements measuring tax attitudes. These items were from the Motivational Postures *Commitment* subscale (Braithwaite, 2003; e.g., "Paying tax is the right thing to do."), which is defined as the moral obligation to pay taxes and to support the principles of taxation.

In the third block, we measured individuals' gender and age, and they were asked to indicate whether they live in the USA.

In the fourth block, we asked participants to list the number of wage and non-wage makers in their household and to estimate their annual personal and household income before taxes. We provided five values for both items based on US personal and household income distributions.

In the fifth block, individuals had to provide answers to two multiple-choice attention checks (i.e., "What was the size of the offered discount?", "What was the purchase product?") with one correct answer. Furthermore, they were presented with one multiple-choice item that served as a manipulation check (i.e., "What type of discount was offered?"), again, where one of five options was correct. For all three items, participants could also choose "I don't remember" as a sixth option.

2.1.2.2. Materials for UK sample. In contrast to the various sales tax rates present in the USA, the UK relies on a 20% value added tax charge on most goods and services. To address this difference, we changed two aspects of the UK scenarios.

First, we included different percentage values for the offered discounts. In the tax-related condition the discount was 20% (which corresponds to the VAT in the UK), and in the tax-unrelated condition the discount was 21% to preserve the ratio from the original US study.

Second, increasing the offered discount rates likely increased the attractiveness of driving 30 minutes to receive the discount. While we didn't expect this to influence the relative difference in attractiveness between tax-related and tax-unrelated discount, the effect could have decreased due to ceiling effects of individuals choosing the discount over the regular price. Therefore, we conducted another pre-test, this time among $N = 51$ UK participants (sampled from Prolific Academic). Results showed that the median driving time was 40 minutes. As the driving times in the US and UK samples were different, it is crucial to only focus on the relative difference between discount type conditions and not the absolute levels when comparing country samples.

Potential moderators were the same as measured in the US sample. One smaller change concerned the items measuring political ideology. Here we changed one item's answering scale to range from *extremely left-wing* to *extremely right-wing* (as opposed to *extremely liberal* vs. *extremely conservative*) and adapted the party affiliation item to cover the UK political landscape. Furthermore, items asking about personal and household income were adapted to match UK income distributions.

2.2. Results

2.2.1. Confirmatory analyses

We conducted the same analysis as reported in the original paper. That is, the proportions of individuals choosing a higher time investment to receive a discount between the two discount type conditions using a χ^2-test. A successful replication was defined as significant χ^2-test results. These proportions are depicted in Fig. 1. The proportion of individuals choosing to accept a longer drive to receive a discount was not different between the tax-related and tax-unrelated discount in the USA (66% vs. 64%), $\chi^2(1, N = 590) = 0.17$, $p = .685$, $r_\Phi = 0.02$, 95% CI [−0.06, 0.10], nor in the UK (80% vs. 79%), $\chi^2(1, N = 595) = 0.05$, $p = .823$, $r_\Phi = 0.01$, 95% CI [−0.07, 0.09].

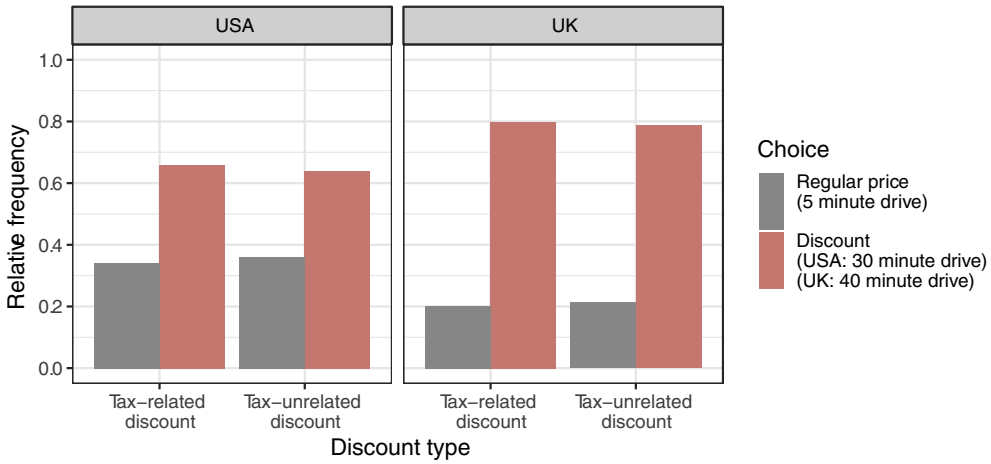

Fig. 1. Proportions of individuals choosing to invest time to receive a discount by discount type and sample.

We regarded the results as informative of a failed replication if the effects were non-significant (χ^2-tests reported above) paired with significantly smaller estimates than an effect detectable with 33% power for $N = 191$; i.e., $r = 0.11$. This comparison assessed whether our replication results are different in magnitude from an effect that would have been detectable with the sample size of the original study (Simonsohn, 2015). Fig. 2 shows the effect estimates of Experiment 1 from the original study and the two replication studies. Both replication sample effects were smaller than the original study's downward corrected effect size (USA: $p = .011$, UK: $p = .006$). Therefore, the replication results of Experiment 1 for both the US and the UK sample were informative of failed replications.

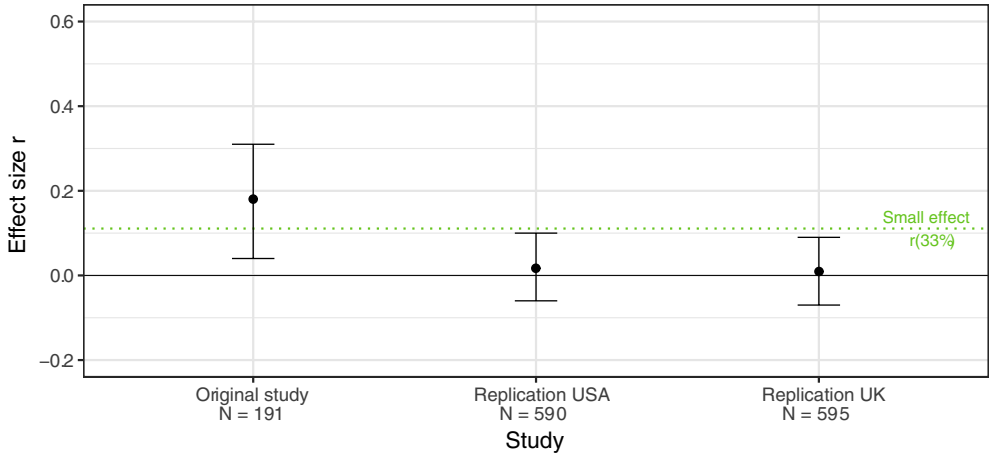

Fig. 2. Results from Sussman and Olivola (2011) Experiment 1 and the two replications. Points indicate effect estimates, and the vertical bars their 95% confidence intervals. The dashed line indicates the effect size that would give the original study 33% power.

To test for differences between the two samples, we ran a logistic regression with decision as the dependent variable (Table 1). Independent variables were the factors discount type, sample, and the interaction of the two variables. The interaction term revealed no choice differences between the two samples with regard to the two discount types. As noted above, the significant simple effect of sample was not of interest.

2.2.2. *Exploratory analyses*

To test whether the effect of discount type on the willingness to drive longer was moderated by a third variable, we measured a number of potentially relevant moderators (recall Section 2.1.2.1), which were suggested during the review process and preregistered. These were political ideology, political party preference,[2] political extremity,[3] political satisfaction, tax attitudes, personal income, household income, attention checks, and a manipulation check. We ran separate regression models with choice as the dependent variable and discount type, a proposed moderator (centered if non-categorical), and the interaction between discount type and moderator as predictors. We ran each model separately for the US and UK sample, resulting in a total of 21 moderation tests.

We found only a single significant moderation, where the decision to drive longer to receive a discount was moderated by tax attitudes in the US sample; Interaction: $B = -0.32$, $p = .047$ (see Table S2 for the full model). The interaction expresses a larger difference between the tax-related and tax-unrelated discount condition for individuals with lower tax attitudes (i.e., those who report that they dislike paying taxes). A closer graphical inspection (Fig. S2) revealed that these individuals seem to dislike tax-unrelated discounts, while they do not seem to differ from individuals with more positive tax attitudes in their preference for tax-related discounts.

Furthermore, we explored whether the main regression model reported in Table 1 was robust after controlling for individuals' demographics and various self-reported measures. Controlling for these variables did not noticeably influence the effect estimate of interest (see Table S3 for details).

[2] We tested this moderator multiple times. The first model only included the two major parties. The second model included all answering options. The third model (US sample only) included the two major parties along with Libertarians.

[3] We created this moderator by "folding over" the political ideology scale, where lower values then expressed more moderate political ideology whereas higher values expressed either being more liberal/left-wing or conservative/right-wing.

Table 1
Logistic regression predicting choice to drive longer in Experiment 1 as a function of discount type, sample, and their interaction.

	Choice to drive longer to receive a discount		
	B	Exp(*B*)	*p*
Intercept	0.57		< .001
Discount type	0.08	1.09	.623
Sample	0.74	2.09	< .001
Discount type × Sample	−0.02	0.98	.943

Note. $N = 1185$. Independent variables were dummy coded: Discount type (0 = tax-unrelated discount, 1 = tax-related discount) and sample (0 = USA, 1 = UK).

3. Experiment 2

3.1. Method

See Table S4 of the supplementary material for the 36-question guide to the Replication Recipe (Brandt et al., 2014) that addresses details about the nature of the original effect and differences between the original and replication study.

3.1.1. Participants

3.1.1.1. Original study. In the original experiment, a total of 401 participants were recruited using MTurk. After excluding 50 individuals (non-US residents, already participated in same study, under the age of 18, short completion time), the final sample was $N = 351$ with 64% female and a mean age of $M = 35.2$ years ($SD = 12.6$).

3.1.1.2. Power analysis. Experiment 2 reports on two different choice scenarios, again comparing individuals who were offered either a tax-related or tax-unrelated discount (between-subject). In the first scenario, individuals were asked how long they would be willing to wait in line in a shopping mall to receive a 9% ("axe-the-tax"/"customer rewards") discount by answering binary yes/no questions for 5-minute time intervals increasing from 5 to 60 minutes. The proportion of Yes responses was higher in the tax-related discount condition (53% vs. 43%), $U = 3.11$, $p = .002$. In the second scenario, participants had to indicate the discount size necessary for them to wait 15 minutes in line. Again, this was measured using binary yes/no choices for discounts ranging from 5% to 12%. The proportion of Yes responses did not differ significantly between the two conditions (72% vs. 79%), $U = 1.81$, $p = .070$, but was interpreted as marginally significant in the original article.

Both scenarios were embedded in a larger questionnaire in a random order. It was therefore unlikely for the two choices to be taken in direct succession. To avoid direct contrast effects of the two choices we decided to limit the replication attempt to the first scenario that yielded a larger effect.

Running a power analysis for a *U* test with $\alpha = 0.05$ and $1 - \beta$ (power) $= 0.95$ for the original effect ($r = 0.15$) yielded a required sample size of $N = 594$.[4] Applying the two downward adjustment rules, safeguard power (60% CI [0.21, 0.39]) and $2.5 \times N$ rule (original $N = 351$), required sample sizes were $N = 1238$ and $N = 878$, respectively. Given the available resources we set the targeted sample size at $N = 700$ for a replication of the first scenario of Experiment 2. This provided 95% power for effects as small as $r = 0.14$ and still 80% power for effects as small as $r = 0.11$ in each sample.

3.1.1.3. Replication study samples and procedure. A total of $N = 702$ US and $N = 699$ UK individuals were recruited through Prolific Academics. The inclusion criteria were being US/UK residents, having provided consistent choices in the list of binary waiting time items (e.g., willingness to wait 15 minutes but not 10 minutes to receive a discount would be inconsistent; see below), and not having participated in our Experiment 1. After exclusions, the final samples were $N = 650$ for the USA and $N = 673$ for the UK. Sample demographics were similar to Experiment 1 with 50% and 62% female participants, and a mean age of $M = 34.6$ years ($SD = 11.5$) and $M = 39.2$ years ($SD = 12.6$) in the USA and UK, respectively. See Fig. S3 for key demographics by sample and condition.

Data collection took place in late November 2018. Median participation time was slightly above three minutes in both samples. Individuals were paid a remuneration of $1.00 for completion.

3.1.2. Materials

3.1.2.1. Materials for US sample. Materials were fully available from the original publication. Furthermore, the original authors were contacted to ask whether there were further unreported procedural details important for the study design, which was not the case.

The study comprised a two-group between-subject design. Both groups read a scenario about a purchase decision for a jacket.

[4] Extracting exact effect sizes from the original paper was not possible, as only *U*-values of non-parametric *U*-tests without corresponding *z*-values were reported. We contacted the authors who shared the raw data. Effect sizes were $r = 0.15$, 95% CI [0.05, 0.25], and $r = 0.10$, 95% CI [−0.01, 0.20], respectively.

Participants were told about the possibility to receive a 9% discount if they wait in line in one of two stores. They had to indicate how long they would be willing to wait to receive the discount. The between-subject factor varied the type of discount ("axe-the-tax" vs. "customer rewards"). The exact scenario stated the following:

> "Imagine that you are walking through the mall looking for a particular jacket that you have seen advertised. You come across two closely located stores that carry it. The first store offers no discounts, but has no wait to purchase the coat. The second store is having a special ["axe-the-tax" sale, with the store selling all items tax-free, equivalent to a 9% discount/"customer rewards" sale, with the store selling all items at a 9% discount]. However, due to the popularity of the sale, there is a wait to purchase items there. How long would you wait in line to receive the discount?"

Following the scenario, individuals had to answer a series of 12 binary yes/no choice questions that asked if they would be willing to wait X minutes to receive the 9% "axe-the- tax"/"customer rewards" savings, with X increasing in 5-minute intervals from 5 to 60 minutes.

In a pre-test we asked $N = 49$ US individuals (sampled from Prolific Academic) how long they would be willing to wait in line to receive a 9% discount by answering binary yes/no questions for 5-minute time intervals increasing from 5 to 100 minutes. We preregistered that we would use the pre-tested 90th percentile waiting time as the maximum value for the binary yes/no questions used after the scenario, which was 50 minutes. Since this time was 60 minutes in a pre-test among $N = 49$ UK individuals, we decided to use the same scale maximum in both samples (i.e., 60 minutes).

After the decision to wait in line, we asked the same post-experimental questions as presented for Experiment 1 (see above).

3.1.2.2. Materials for UK sample. As in Experiment 1, materials were adapted to suit the 20% value added tax used in the UK. That is, the discount size in the scenario was set to 20%. Another minor change addressed language differences. We replaced the word "mall" with "shopping centre" which is more appropriate for a UK sample. All remaining aspects of the materials corresponded to the US version.

3.2. Results

3.2.1. Confirmatory analyses

Again, the analysis was conducted in line with the original paper. First, we calculated an average score of Yes responses for each individual, indicating every individual's willingness to wait. We compared the Yes proportions between the two discount groups ("axe-the-tax"/"customer rewards") using a *U* test for both samples separately. A successful replication was defined as a significant *U* test. The relative number of Yes choices by time interval are depicted in Fig. 3. The proportion of Yes choices was significantly higher in the tax-related as compared to the tax-unrelated discount condition in the USA (46% vs. 41%), $W = 58789$, $z = -2.27$, $p = .012$, $r = 0.09$, 95% CI [0.01, 0.16], making the replication successful in this sample. However, the effect was not observed in the UK (56% vs. 53%), $W = 59804$, $z = -0.85$, $p = .197$, $r = 0.03$, 95% CI [−0.04, 0.11].

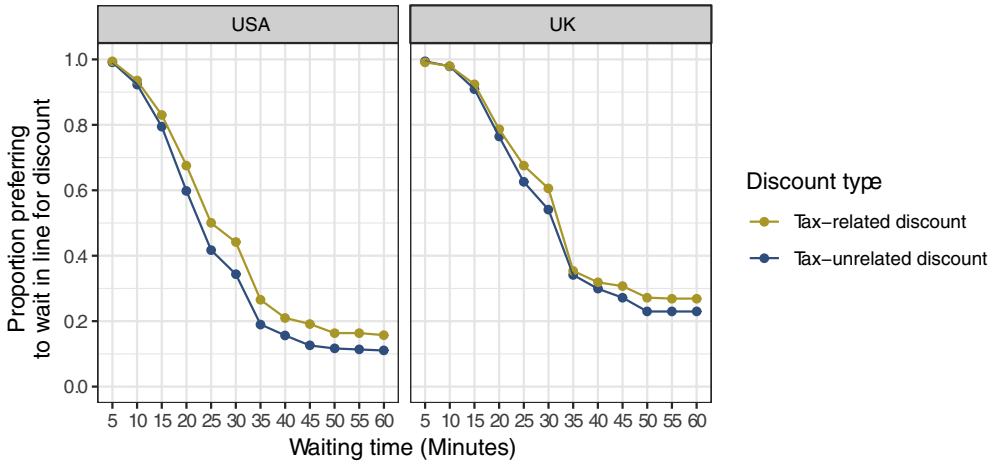

Fig. 3. Proportions of individuals willing to wait in line for each time interval by discount type and sample.

We also tested whether the replication results were different from an effect size that is large enough to still have been detectable (i.e., with 33% power) with the sample size of the original study (Simonsohn, 2015). We regarded replication failure as informative if we failed to find a significant *U* test along with an effect estimate significantly smaller than an effect detectable with 33% power for $N = 351$; i.e., $r = 0.08$. Fig. 4 shows the effect estimates of Experiment 2 from the original study and the two replication studies. Given the successful replication in the USA and the size of the original effect size, one would expect that the US replication effect is not different from an effect size that is plausible for the original sample size, which was the case ($p = .426$). The UK effect, which was not different from zero, was also not different from $r = 0.08$ ($p = .103$). Therefore, replication results of Experiment 2 were successful in the USA but inconclusive in the UK.

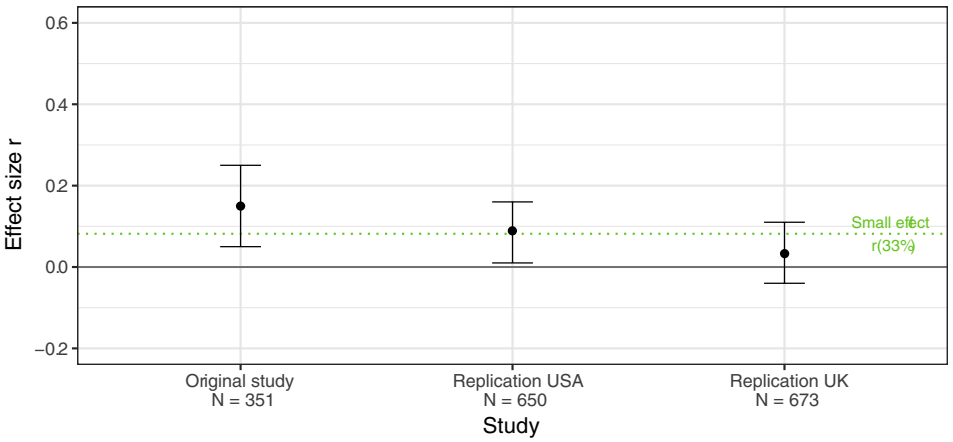

Fig. 4. Results from Sussman and Olivola (2011) Experiment 2 and the two replications. Points indicate effect estimates, and the vertical bars their 95% confidence intervals. The dashed line indicates the effect size that would give the original study 33% power.

To test for differences between the two samples, we ran a linear regression with waiting time (proportion of Yes choices; Table 2)[5] predicted by discount type, sample, and the interaction of the two variables. Results suggested a simple effect of discount type, expressing an increased willingness to wait in line in the tax-related condition in the US sample. While the interaction term was negative, pointing towards an effect mitigation in the UK sample, the interaction was not significant. Hence, the regression attested an overall effect for the full sample.

Table 2
Linear regression predicting willingness to wait in line in Experiment 2 as a function of discount type, sample, and their interaction.

	Willingness to wait to receive a discount		
	B	β	*p*
Intercept	0.41		< .001
Discount type	0.05	0.09	.016
Sample	0.13	0.22	< .001
Discount type × Sample	−0.03	−0.04	.409

Note. $N = 1323$. Independent variables were dummy coded: discount type (0 = tax-unrelated discount, 1 = tax-related discount) and sample (0 = USA, 1 = UK).

3.2.2. *Exploratory analyses*

We conducted the same preregistered exploratory analyses as in Experiment 1. None of the interactions between discount type and a potential moderator was significant in either sample.

Furthermore, we explored whether the main regression model reported in Table 2 was robust after controlling for individuals' demographics and various self-reported measures. See Table S5 for two additional models. Controlling for demographics and various self-reported measures did not noticeably influence the results.

4. Discussion

In the original study, Sussman and Olivola (2011) demonstrated that US individuals are more willing to incur a cost (i.e., longer drive to a store or waiting in line) to avoid paying taxes than to avoid other equivalent costs that are tax-unrelated. The results suggest that people's dislike of taxes influences everyday decisions beyond the immediate decision context of taxes.

Our replication study consisted of replications of Experiment 1 and 2 of the original study. Close replications in high-powered US samples only provided partial support for this effect. Specifically, while we could not replicate the effect in Experiment 1, in Experiment 2 we found a small effect. We conclude that Tax Aversion, defined as a dislike of taxes per se, measured in the context of hypothetical purchase decisions, is observable in the USA, but only constitutes a small effect.

The salience of consumption taxes is clearly larger in a sales tax system, where taxes constitute an immediate loss added at the

[5] In the stage 1 registered report we stated we would run a mixed effects regression here. Comparing a mixed model against a more simplified model applied here revealed no interpretational differences and only minor *p*-value deviations for all calculated models. We therefore consistently report the simpler models here. The initially stated analyses are included in the R code uploaded to the main OSF page.

check-out, as compared to a price-inclusive value added tax system (Bird, 2010). It seems reasonable to assume that such a salience, at least in part, may contribute to the observed effect in US samples. To test this assumption, we extended the replication efforts to the UK, a country employing value added tax. In combining results from both experiments, we found no substantial support for Tax Aversion in the UK. The observed country differences support the notion that the salience of a tax may influences its evaluation. An alternative explanation could be that sales strategies using tax-free labels might be less common in the UK than in the USA.

The results suggest that Tax Aversion is a less pronounced phenomenon in the USA than originally proposed by Sussman and Olivola (2011). Effect size point estimates for the two experiments dropped from $r = 0.18$ and $r = 0.15$ reported in the original study down to $r = 0.02$ and $r = 0.09$ in our close replications. Assuming a small effect exists in the US sampled population, the overall inconclusive results in terms of differences between Experiment 1 (absence of an effect) and Experiment 2 (detection of a small effect) could be explained by the dependent measures used. Experiment 1 relied on a simple binary choice, forcing individuals to decide for one of two positions, whereas Experiment 2 was able to detect more subtle differences in preferences.

In Experiment 4 of the original paper (Sussman & Olivola, 2011), participants affiliated with antitax parties showed stronger Tax Aversion both in the USA and UK. This is consistent with previous research linking tax attitudes to political ideology. For instance, left leaning voters often report more positive tax attitudes than right leaning voters (Olsen, Kogler, Stark, & Kirchler, 2017; Wahlund, 1992). Political satisfaction with the current political legislation has also been found to be positively associated with individuals' attitudes toward paying taxes (Svallfors, 2013). However, after exploring various constructs from the political domain as potential moderators (i.e., political ideology, party preference, extremity, satisfaction), we found no moderation effects, in either experiment or sample.

An alternative explanation for the lack of significant moderation effects could be poor data quality. However, we think this is unlikely as we find expectable and meaningful correlations between measured tax attitudes and political ideology (i.e., more positive tax attitudes among individuals considering themselves more liberal or left-wing, and Democrat or Labour [the latter only in Experiment 1]), suggesting that participants display consistent preferences between these.

Furthermore, attention and manipulation check success rates, combined with no moderation effects of check item correctness, increase the confidence we have in the data quality. For Experiment 1 and 2, respectively, the discount size was correctly recalled by 95% and 99% of individuals, and the product in 96% and 90% of cases. The discount type (i.e., manipulation check) was correctly recalled in 86% and 90% of cases, while these numbers were even higher in the tax-related conditions with 92% and 99%. There were no sample differences.

We do *not* claim that the spillover of tax attitudes on everyday decisions must generally be absent or small. We assume that tax attitudes are likely to influence behavior if true consequences can be expected (e.g., political referenda). In reality, however, tax-free advertising is just a sales strategy without any tax revenue consequences. Sales taxes still apply (the stores still forward consumption taxes to the authorities). Participants might have been aware of this, especially those who are tax averse and would willingly avoid tax payments if possible. Future studies on Tax Aversion might want to take this aspect into account and study the phenomenon in a decision context where taxes can truly be avoided.

Author note

Main OSF page (data, R code, materials): https://osf.io/q8g7f/; Stage 1 registered report: https://osf.io/7pruq/; Preregistration of pre-tests: https://osf.io/g2jbr/; Preregistration of exploratory analyses: https://osf.io/nqg3e/.

Acknowledgment

We thank Abigail Sussman and Christopher Olivola for supporting our replication attempt by providing data, confirmation of materials, and important feedback. We also thank Žiga Puklavec for helpful comments on previous versions of this manuscript and Marco Rapp for assisting in a bibliometric search. Linda Dezső is grateful for the Back to Research Grant from the University of Vienna for funding work on this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joep.2019.02.001.

References

Bird, R. M. (2010). Visibility and accountability: Is tax-inclusive pricing a good thing? *Canadian Tax Journal, 58*, 63–76.

Braithwaite, V. (2003). Taxing democracy: Understanding tax avoidance and evasion. In V. Braithwaite (Ed.). *Dancing with tax authorities: Motivational postures and non-compliant actions* (pp. 15–39). Aldershot, U.K.: Ashgate.

Bramley, G., Grand, J. Le, & Low, W. (1989). How far is the poll tax a "community charge"? The implications of service usage evidence. *Policy and Politics, 17*, 187–205.

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-sorolla, R., & van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review, 99*, 1145–1177. https://doi.org/10.1257/aer.99.4.1145.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010). A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psychological Science, 21*,

86–92. https://doi.org/10.1177/0956797609355572.
Mccaffery, E. J., & Baron, J. (2006). Thinking about tax. *Psychology, Public Policy, and Law, 12*, 106–135. https://doi.org/10.1037/1076-8971.12.106.106.
Olsen, J., Kogler, C., Stark, J., & Kirchler, E. (2017). Income tax versus value added tax: A mixed-methods comparison of social representations. *Journal of Tax Administration, 3*, 87–107.
Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science, 9*, 319–332. https://doi.org/10.1177/1745691614528519.
Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*, 559–569. https://doi.org/10.1177/0956797614567341.
Smith, P. (1991). Lessons from the British poll tax disaster. *National Tax Journal, 44*, 431–436.
Sussman, A. B., & Olivola, C. Y. (2011). Axe the tax: Taxes are disliked more than equivalent costs. *Journal of Marketing Research, 48*, 91–101. https://doi.org/10.1509/jmkr.48.SPL.S91.
Svallfors, S. (2013). Government quality, egalitarianism, and attitudes to taxes and social spending: A European comparison. *European Political Science Review, 5*, 363–380. https://doi.org/10.1017/S175577391200015X.
Wahlund, R. (1992). Tax changes and economic behavior: The case of tax evasion. *Journal of Economic Psychology, 13*, 657–677. https://doi.org/10.1016/0167-4870(92)90017-2.
Winetrobe, B. Y. B. K. (1992). A tax by any other name: The poll tax and the community charge. *Parliamentary Affairs, 45*, 420–427. https://doi.org/10.1093/oxfordjournals.pa.a052369.