Antoine Genest-Grégoire[1]

Carleton University, School of Public Policy and Administration

Luc Godbout

Université de Sherbrooke, Department of Taxation

Jean-Herman Guay

Université de Sherbrooke, School of Applied Political Studies

[1] Corresponding author, a.genestgregoire@gmail.com

This research has been made possible by the ongoing financial support of Québec's Ministry of Finance to the activities of the Research Chair in Taxation and Public Finance at Université de Sherbrooke. Antoine Genest-Grégoire also aknowledges the support of the Fonds de recherche du Québec - Société et Culture and the government of Canada's Vanier-Banting Secretariat. The authors would also like to thank David Boisclair, Kevin Milligan, Jennifer Robson, Frances Woolley, Olivier Jacques and Gregory Eady for helpful comments on the original draft of the survey questionnaire and this article.

Lists: a novel experimental method to measure tax evasion

Abstract

For governments trying to design interventions to raise tax compliance, better understanding of the individual and social roots of illegal actions to reduce payments is very valuable. However, the validity of such behavioural models and interventions cannot be assessed without tackling the question of tax evasion measurement. This paper presents a novel experimental method to measure the propensity to evade taxes that overcomes sensitivity bias. By using list experiments, it provides estimates of the prevalence of tax evasion amongst different groups of the population. Using this method from political science research, we find that 13 percent of the Canadian population admit to evading income taxes, while 29 percent admit to evading consumption taxes. These estimates are robust to various changes in model specification and they are of a comparable magnitude to estimates made using audit data (Kleven et al., 2011). The estimated propensities are lower for women or older respondents. Higher income is associated with less income-tax evasion and more consumption-tax evasion, as is being self-employed. Our results provide a test to the usefulness of list experiments to measure both income and consumption-tax evasion (Fergusson et al. 2019). The technique should provide researchers with a method to test recent theories about individual causes of evasion such as tax morale (Luttmer and Singhall, 2014) or trust in authority (Kirchler et al. 2008).

Keywords:

Tax, Tax Evasion, Tax Compliance, List Experiments, Surveys, Experimental methods, Social Desirability Bias

Introduction

Tax evasion is a difficult subject to study. Even when a definition is agreed upon by different groups, measurement of the phenomenon remains extremely difficult, because of its immoral character. Researchers can expect that any source of information on the subject will suffer from some type of sensitivity bias such as social desirability, fear of response disclosure or intrusiveness (Blair, Imai and Coppock, 2020). This essay is focused on the question of measurement and defines tax evasion as "illegal and intentional actions taken by individuals to reduce their legally due tax obligations" (Alm, 2012).

Thinking about how much revenue is lost to evasion hides two questions: who evades taxes and by how much? Our research is mostly focused on this second question. Macro research on the question aggregates the two but a growing literature, inspired by psychology and behavioural economics, tries to answer specifically the first question. Essentially, research on the subject is progressing in a pincer movement: macroeconomic data allows us to estimate the total scope of the phenomenon and micro theory tries to explain the individual drivers of evasion (Slemrod and Weber, 2012). This paper aims to contribute to this literature by testing a method to measure tax evasion prevalence, which sits somewhere in the middle. Researchers build models to explain why some taxpayers break the law and others don't. Our experiment seeks to give them an outcome measure to test their theories against rather than the intention or opinion variables they mostly use. Knowing who evades taxes is crucial in understanding how these individuals differ from those that are tax compliant.

Political scientists deal with subjects for which honest answers from citizens are difficult to get and we borrowed one of their methods. List experiments have been used to measure support for same-sez marriage (Luks and Monson, 2010), attitudes towards race ((Kuklinski, Cobb, Gilens, 1997) and electoral turnout (Holbrook and Krosnick, 2010), amongst many other subjects. The tool belongs to the family of indirect questioning techniques. The general idea is that if the subject is sensitive, respondents might be more inclined to give truthful answers if they feel their privacy is protected. List experiments hide the opinion of each individual respondent on the subject, while still allowing the researcher to measure it over the whole respondent sample.

This paper includes an overview of the research literature on tax evasion measurement and the list experiment method, a presentation of headline results from the experiment, alternative specifications and tests of the robustness of these results, a brief discussion, and a conclusion.

Literature

This paper brings a method mostly used in political science to bear on a problem mostly studied by economists. To our knowledge, only one research paper (Fergusson, Molina and Riano 2019) has pursued this new method-subject association so we feel an overview of both tax evasion and list experiment literature is warranted. We first describe the state of research on tax evasion, to situate our results, and we then describe what list experiments are and how they can be useful for the study of such a subject.

Tax Evasion

The seminal paper on tax evasion was written in 1972 by Allingham and Sandmo. They essentially describe tax evasion as a gamble taxpayers take, based on their tax liability, the odds of their return being audited and the penalty if they are found to have cheated. Central to this model is the idea of expected utility: individuals have preferences for different outcomes and for certainty or uncertainty that combine to explain how they make choices when confronted with uncertain outcomes. While research has confirmed some intuitions of this model (Clotfelter 1983, Andreoni, Erard and Feinstein, 1998, for an overview), a pure expected utility model cannot explain the actual level of tax evasion we see in the world (Alm, McClelland, and Schulze, 1992, Dwenger, Kleven, Rasul and Rincke, 2016). Indeed, audit rates and penalty levels are way too low to explain the level of compliance we observe.

For this reason, extensions to this model have included either bounded rationality or social interactions (Alm, 2012). The first type of model seeks to explain why taxpayers fear audits or fines to a level that does not seem commensurate with their frequency or bite by studying how we perceive information and make decisions (Hashimzade, Myles and Tran-Nam, 2013). The second strand of research seeks to explain tax compliance by looking at how taxpayers' behaviour is shaped by their interactions with others (Luttmer and Singhal, 2014). This can take many forms such as reference groups, peer pressure or patriotism.

Measurement is a fundamental difficulty when studying an illegal phenomenon. Traditional estimations of the extent of tax evasion are macroeconomic and they measure the total loss of tax revenue per tax base (Internal Revenue Service, 1996, Slemrod, 2007). For example, the Canadian Revenue Agency (CRA) produced estimates that place the tax gap at 6.4 percent of revenue for the Canadian federal income tax (CRA 2017) and 5.6 percent of value-added tax revenue (CRA 2016). As our models explaining tax evasion become more sophisticated, more fine-grained methods of data collection become necessary. Moving from estimations of tax evasion that depend on tax levels or audit rates to estimations based on perceptions of detection probability or perceived social norms require measures at the individual level to test theory (Gemmell and Hasseldine, 2012). Researchers such as Kleven and his colleagues (2011) have produced individual tax evasion estimates using administrative tax data (10,7 % of taxpayers evade the Danish income tax), but they lack information about taxpayers that would help to understand their behaviour. This explains the rise of experiments in the study of tax evasion (Coricelli, Joffily, Montmarquette and Villeval, 2010, Fortin, Lacroix and Villeval, 2007). Not only do we want to know how many individuals add up to the total revenue loss, we want to know who these individuals are.

List Experiments

List experiments rely on asking a sensitive question indirectly. Rather than asking respondents if they hold a sensitive attitude or if they have taken part in a sensitive activity, respondents are asked the number of statements about behaviour or attitudes that applies to them out of a list. Half of the respondents see a list that contains only control items that are of no interest to the researcher. The other half sees a list with the same control items plus the sensitive item of interest. If the experiment was executed properly, the difference in the mean number of items to which respondents agree provides an estimate of the prevalence of the sensitive item of interest.

There are multiple methods to indirectly obtain answers to sensitive questions. Amongst them are the item count technique, of which the list experiment is the most common application, the randomized response technique, and the endorsement experiment. All these methods have in common that they offer respondents a way to hide their answer either from the researchers or from their peers, as part of the questionnaire design (Tourangeau and Yan, 2007, Chou, Imai and Rosenfeld 2017). A vast literature documents the difficulty of getting truthful answers on sensitive questions and indirect measurement methods have had some success in alleviating this difficulty (Coutts and Jan 2011, Rosenfeld, Imai and Shapiro, 2016). Heerwig and McCabe's (2009) research on support for a black presidential candidate is perhaps the most famous example of the use of the list experiment, but Blair, Coppock, and Moor (forthcoming) found 147 published research papers using one variant or the other of the list experiment[2].

[2] The list experiment is the most popular of these three indirect methods. The randomized response technique combines the respondent's numerical answer to a question of interest with a random number, such as their birthday or the result of a dice throw. The researcher cannot know each respondent's specific answer but since she knows the distribution of the random variable, she can estimate the distribution of the underlying number if the sample is large enough. The endorsement experiment works by asking respondents their support for a group (or person). Half the respondents see the basic version of this question while the other is primed by learning that the group is associated with the item of interest for the researcher. The difference in approval for that same group provides an estimate for the item of interest's approval (eg. "Montrealers are known to evade

Since its inception, the method has been refined by including double-list experiments (Droitcour et al. 1991), by selecting control items with negative valence (Glynn, 2013) and low and high prevalence (Kuklinski, Cobb, Gilens, 1997), by combining it with endorsement experiments (Blair, Imai and Lyall, 2014), or with direct questions (Aronow, Coppock, Crawford and Green, 2015).

Researchers have also refined the statistical analysis of list experiments in ways that allow for estimations of the link between individual characteristics and the measured trait or attitude (Imai, 2011, Blair and Imai 2012) and make it more robust to violations of some of its assumptions (Blair, Chou and Imai, 2019). Such multivariate methods use the information obtained from respondents who agree with all or none of the items presented as part of the list to create a joint distribution of the number of answers to control and sensitive items (shown in Table 4) and then estimate the association between this distribution and regressors of interest included in the survey. This produces a binomial logistic regression model with large variance but interpretable coefficients, once transformed into marginal effects (see Table 3). We based our estimations on the maximum likelihood estimation method described by Imai (2011).

Methods

Our list experiment used four control items. Following Glynn (2013), our groups of control items included two elements with an expected negative covariance, one element with high expected prevalence and one with very low expected prevalence. These elements contribute to enhancing respondent confidentiality: high and low prevalence items lower the odds of respondents answering 5 or 0, which deprives them of confidentiality. If a respondent answers 0 out of 5, then we know with certainty that she did not evade taxes and if she answers 5 out of 5 we know that she did with certainty. Having two items that are negatively correlated protects confidentiality because even if the odds of agreeing to one of the two are very high, researchers can't know which one applies to each respondent[3]. Our survey also included questions asking respondents directly whether they evaded taxes or not at the end of the questionnaire and after they had been exposed to the experiment (Aronow et al. 2015 details how to use the combination of direct questions and lists). The list experiment is useful if respondents feel like the item

taxes more than other Canadians."(prime), "What is your opinion of Montrealers?" (endorsement question)).

[3] Because our survey was framed as concerning taxation and governments, we tried to have control items that were political in nature and somewhat related to tax. We choose typical left-wing and right-wing propositions as inversely correlated items. The items were slightly different from those used by other researchers because politics in Canada is defined along both the left-right and federalist-nationalist axis. See Lax et al., 2016, and Ahlquist, 2018, for more reflections about the selection of control items.

of interest is sensitive. Asking a direct question about the subject to all respondents is a way of validating that respondents indeed find the subject sensitive.

We used two list experiments, one to measure the evasion of income taxes and the other to measure the evasion of consumption taxes[4]. Figures 1 and 2 summarize these experiments.

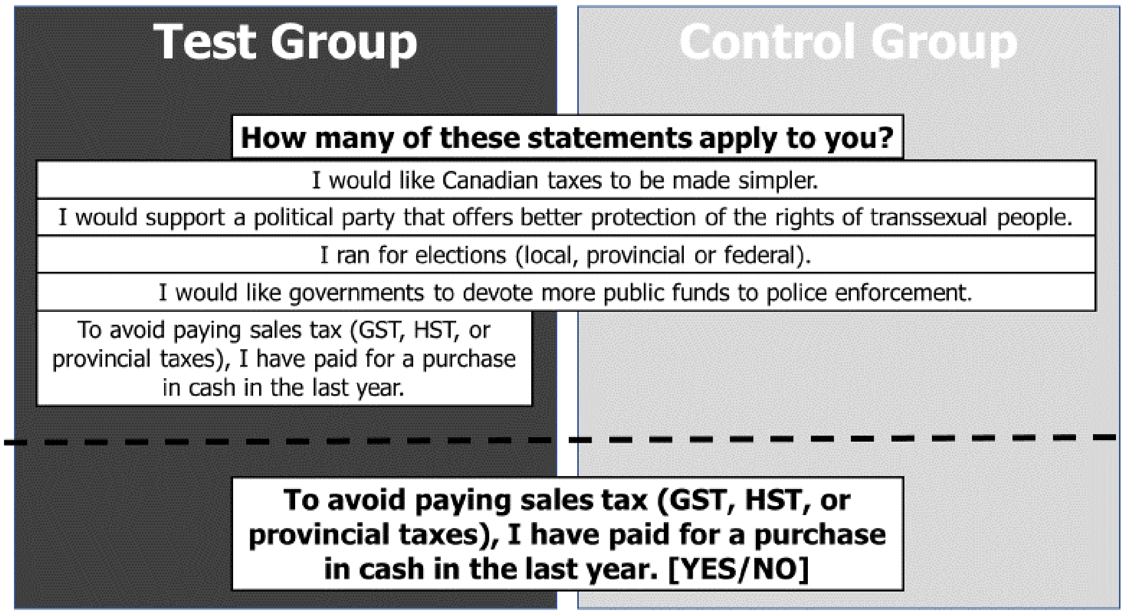

Figure 1: Consumption-Tax Evasion List and Direct Question

[4] Canadians pay a federal value-added tax (the Goods and Services Tax or GST) and some also pay provincial sales taxes (PST) or provincial value-added taxes assessed under the GST framework (Harmonized Sales Tax or HST). Since we measured compliance to these taxes from the point of view of the consumer, they were all lumped in together.

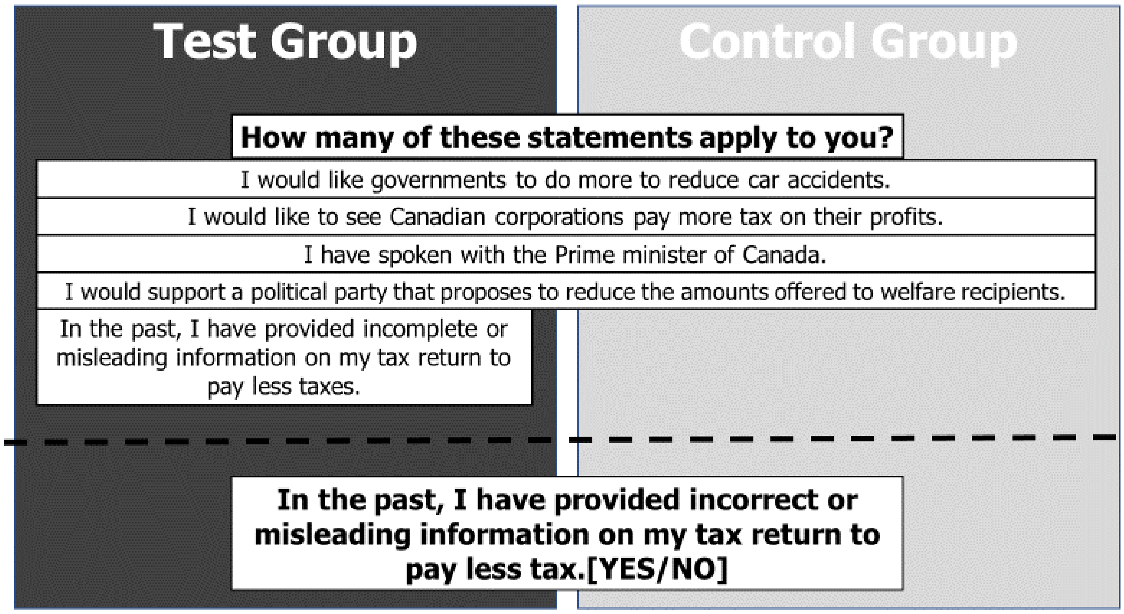

Figure 2: Income-Tax Evasion List and Direct Question

Respondents could be part of the test or control group for each of these two experiments. The order of appearance of the two lists was randomized, creating 8 possible paths for respondents (see figure 3). For example, a respondent on path 5 could have answered that she agreed to 4 items out of 5 on the first list she was presented and 2 out of 4 in the second.

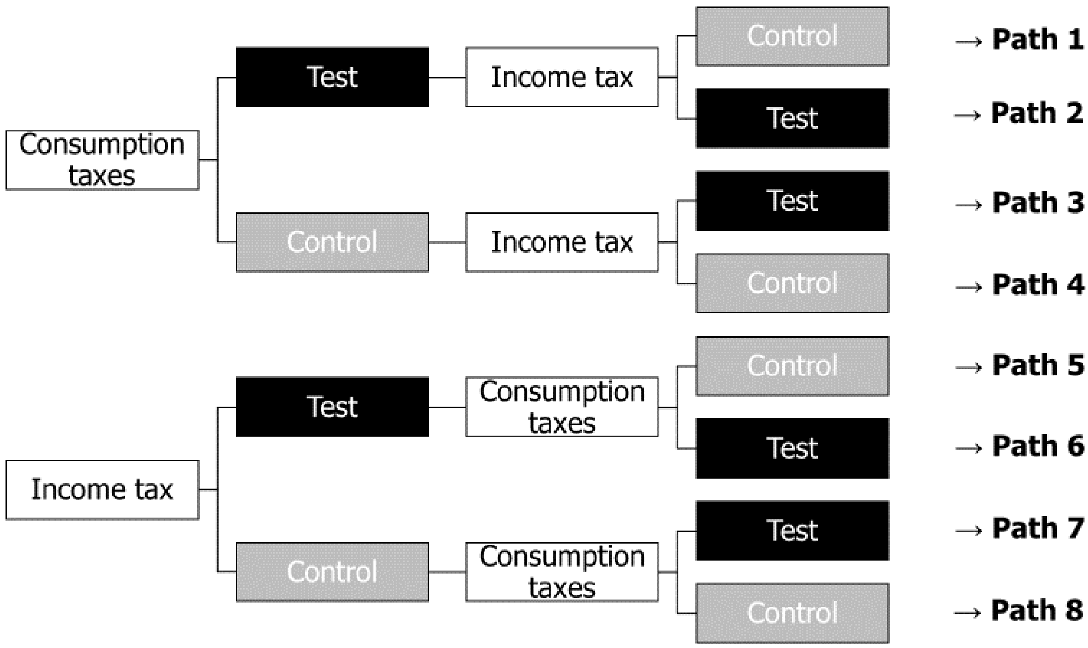

Figure 3: Eight possible Experiment Paths

The survey also included other questions about attitudes and knowledge about taxes. These questions were placed between the two lists and between the lists and the direct question, to prevent respondents from detecting that the subject of tax evasion was of specific interest to designers of the questionnaire, as recommended by Eady (2017).

We implemented our list experiment as a survey, which helped us reach the large sample size required to compensate for the increased variance associated with list experiments compared to direct questions. Our online-based survey had 2806 respondents after data cleanup[5] (response rate of 14% according to AAPOR (2016) RR1 definition). The sample was stratified by province, age and gender, education and income to ensure comparable groups in all branches of the experiment[6][7]. The survey was administered by polling firm Léger which uses a Web panel. The panel includes 420,000 Canadians, the majority of its members have been selected by random phone dialling and the rest by third-party vendors or through social media. The survey was pretested with 35 respondents.

Table 1: Sample description

Characteristic	N	%	Characteristic	N	%
Men	1496	53.3	**18 to 24 years old**	284	10.1
Women	1310	46.7	**25 to 34 years old**	475	16.9
Full-time employee	1250	44.5	**35 to 44 years old**	474	16.9
Part-time employee	261	9.3	**45 to 54 years old**	566	20.2
Self-employed	177	6.3	**55 to 64 years old**	481	17.1
Student	196	7	**65 to 74 years old**	388	13.8
At home	110	3.9	**75 or older**	138	4.9
Unemployed	160	5.7	**Family income below $20,000**	254	9.1
Retired	635	22.6	**20 to $40,000**	478	17
No answer	17	0.6	**40 to $60,000**	516	18.4
Canadian Native	2332	83.1	**60 to $80,000**	453	16.1
Non-native to Canada	474	16.9	**80 to $100,000**	401	14.3
			$100,000 and above	704	25.1
			High School Diploma or less	715	25.5
			College Diploma	863	30.8

[5] The original dataset includes 3156 respondents. Respondents who did not answer any of the questions used in the analysis, such as the one asking their family income, were removed.
[6] We were provided survey weights by the polling company but chose not to use them for the analysis as some of the statistical routines we use are not adapted to weighting yet.
[7] See Appendix A for the distribution of respondents' characteristics by subsample.

			University Diploma	1228	43.8

Results

Table 2: Prevalence Estimates of Tax Evasion

	Number of items—control	**Number of items—test**	**Difference**[8]	**Prevalence estimation—list (percent)**	**Prevalence estimation—direct (percent)**
Consumption-tax	1.65	1.93	0.28 ***	28	26.19
Income-tax	1.58	1.71	0.13 ***	13	5.63

The prevalence of consumption-tax evasion is much higher than the prevalence of income-tax evasion, as should be expected given that most income tax is withheld by employers and thus difficult to evade. The difference between direct and list-based estimates is much higher for income-tax evasion (13.5 vs 5.6, p-value<0.000) than for consumption-tax (28.5 vs 26.2, p-value=0.005), even more so in relative terms. These results fit with our intuition that income tax evasion is much more frowned upon than consumption-tax evasion and the comparison might even understate the difference in prevalence as the income tax item concerned evasion at any point in the lifetime of the respondent while the item on consumption taxes asked only about the behaviour in the last year.

Using Blair and Imai's (2012) methods, we estimate regressions for both types of evasion. Our survey included information on gender, age and family income of respondents. We also knew whether they were self-employed and if they were born in Canada or not. We also included provinces as regressors, to capture possible regional effects associated with different tax or economic systems.

[8] Throughout the paper:

*	95% significance
**	99% significance
***	99.9% significance

Table 3: Average Marginal Effects of Individual Characteristics on Propensity to Evade

	Consumption taxes	Income tax
Respondent is		
Female (male)	-18.3 ***	-9.37 ***
Self-employed (any other form of employment)	9.53 ***	-14.39 ***
Non-native to Canada (native to Canada)	0.13	14.86 ***
Age (44 to 55 years old)	***	***
18 to 24 years old	9.68	-2.01
25 to 34 years old	12.53	5.04
35 to 44 years old	-6.85	-0.44
55 to 64 years old	-5.17	1.39
65 to 74 years old	-12.6	-2.3
75 or older	16.98	1.56
Total family income ($100,000 and above**)**	***	***
Less than $20,000	0.88	-1.23
$20,000 to $40,000	1.78	-2.4
$40,000 to $60,000	2.69	-3.5
$60,000 to $80,000	3.62	-4.53
$80,000 to $100,000	4.56	-5.5
Education (university diploma)	***	***
High-school diploma	-3.51	-0.49
College diploma	-17.88	-3.06
Tax filing (tax professional)	***	***
Pen and paper	-11.4	4.2
Tax software	-20.65	6.07
Friend or relative	-12.03	-0.31
Did not file	-30.34	2.05
Province (Ontario)	***	***
Alberta	-9.24	-16.24
British-Colombia	19.96	-4.07

Manitoba	3.46	-1.21
Atlantic	-3.34	-9.04
Quebec	10.44	0.87
Saskatchewan	5.46	-18.5

Note: reference category in parentheses, significance of variables is assessed through an Anova test for multiple category variables (shown next to the category header) and a two-sided t test for dichotomous variables (shown next to the estimate).

Robustness checks

List experiments present risks for statistical validity. Compared to direct questions, list experiments theoretically reduce bias at the expense of variance. If designed correctly, they mitigate sensitivity bias due to the nature of the question. Because they rely on indirect information, much larger samples are required to reach the same statistical power as a direct question. Such experiments lower the risk of not finding an effect because respondents might not want to disclose information but they also raise the odds of not finding such an effect because of increased noise.

List experiments rely on two important hypotheses: no design effect and no liars. The first implies that the number of control items to which respondents agree does not change if they are presented with the additional controversial item. This hypothesis is crucial for the difference in mean number of items between test and control group to be an adequate measure of the prevalence of the tested item. The second hypothesis implies that respondents found their privacy to be adequately protected by the design of the experiment and thus that they gave the true number of items with which they agree. This second hypothesis is implicit in any survey and it is the fact that we don't think it holds when asking respondents about sensitive questions that pushes us to use an indirect method such as lists.

Blair and Imai (2012) have developed a test of the first hypothesis that relies on the number of respondents for each possible answer to the question about the number of items of agreement. If the hypothesis holds, one can calculate the shares of tested respondents who agree with the controversial item for every possible answer to the count question, by using values at the top and bottom of the distribution. Respondents in the test group who say all items apply to them can be assumed to have the controversial trait. If the no design effect hypothesis holds and the treatment and control groups are of the same size, the same number of respondents in the control group have the controversial trait and agree with all 4 control items. Since the number of respondents in the control group saying that 4 items apply to them is known, one can subtract from it to obtain the number of them for whom the controversial trait does not apply. Following this logic, one can estimate the distribution of respondents who agree with 0,1,2,3 or 4 control items and have the controversial trait or not[9]. If these estimates are negative, then the hypothesis does not hold.

Table 4: Estimated Proportions of Tax Evaders by number of Items on Lists

Number of control items	Consumption-tax		Income-tax	
	Percent of respondents who disagree with evasion item	**Percent of respondents who agree with evasion item**	**Percent of respondents who disagree with evasion item**	**Percent of respondents who agree with evasion item**
0	9.84	0.43	11.57	0.38
1	26.16	10.26	32.05	4.08
2	23.59	10.69	32.28	4.92
3	10.91	5.27	9.65	2.01
4	1.00	1.85	0.92	2.14
Total:	71.5	28.5	86.47	13.53

None of the estimated proportion is negative and thus the null hypothesis of no design effect cannot be rejected[10].

Another method to strengthen estimates of prevalence under list experiments is to combine the results of the list with direct questions. This method relies on the hypothesis that no one would lie in the affirmative to a direct question about a controversial idea or behaviour. Thus, the number of respondents who agree with the statement when asked directly provides a floor of support for the statement that can be used to refine estimates from the list experiment (Aronow et al. 2015). It also relies on a stronger version of the no design effect hypothesis: not only must including the controversial item have no effect on the number of control items that the respondents agree with, it must also have no effect on their answer to the direct question. Both these hypotheses can be tested to see if the combined estimation method might produce valid results. On average, respondents who said they did evade taxes when asked directly and were part of the treatment group should agree with 1 more list item than respondents who were part of the control group and didn’t confess evasion when asked directly (so we test whether we can reject the null hypothesis that the difference denoted Beta is equal to 1). Also, the share of respondents confessing to tax evasion when asked directly should be the same

[9] See Blair and Imai’s Table 1 for a good illustration.

[10] Using Blair and Imai (2012) proposed test for the violation of the no-design hypothesis yields p-values of 1 in both cases, much higher than the threshold of alpha/2 = 0.025 for a 5 percent significance level.

in both treatment and control groups (so we test for the rejection of the null that delta=0).

Table 5: Combined Direct-Indirect Method - Consumption-tax Evasion Prevalence Estimate and Hypothesis Tests (n=2806)

	Estimation	**Standard deviation**	**p-value**
Prevalence Estimate (percent)	33.76	3.5	NA
Test 1: H0: Beta=1	0.75	0.07	0.00
Test 2: H0: Delta=0	0.04	0.02	0.01

Table 6: Combined Direct-Indirect Method – Income Tax Evasion Prevalence Estimate and Hypothesis Tests (n=2806)

	Estimation	**Standard deviation**	**p-value**
Prevalence Estimate (percent)	14.81	3.69	NA
Test 1: H0: Beta=1	0.83	0.17	0.32
Test 2: H0: Delta=0	-0.01	0.01	0.53

Tables 4 and 5 show that this first hypothesis is likely violated for the list on consumption-tax evasion but might hold for the one on income-tax evasion (in both tests the rejection of the null implies the violation of the assumption, so large p-values are desirable). The estimate for income-tax evasion shows a prevalence that is significantly higher than the one obtained with the list only (14.81 vs 13.53, $p<0.000$) but that stays in the teens.

The double list design of our experiment might be to blame for the inconclusive results regarding consumption tax evasion. Some respondents were part of a control group for one list and a treatment group for the other. These respondents might have guessed the way our experiment works or, at least, realized that the question of tax evasion was what most interested us. This might have impacted their answer to the direct questions asked later in the questionnaire. To alleviate this problem, we computed the estimates combining list and direct questions for a subsample of respondents who were either in control groups all the time (Paths 4 and 8 in Figure 3) or in treatment groups all the time (Paths 2 and 6). To be even more conservative, we only include respondents who saw the list of interest first amongst the two lists in our test group. We thus compare individuals on paths 4 and 8 with individuals on Path 2 to produce our estimate for consumption taxes and individuals on paths 4 and 8 with those on Path 6 to produce our estimate for income tax evasion[11]. These respondents had the least chance to understand the way the varying number of items in lists could be used to measure behaviour[12].

Table 7: Combined Direct-Indirect Method - Consumption-tax Evasion Prevalence Estimate and Hypothesis Tests – Restricted Sample (N=1054)

	Estimation	**Standard deviation**	**p-value**
Prevalence Estimate (percent)	26.56	6.26	NA
Test 1: H0: Beta=1	0.8	0.13	0.13
Test 2: H0: Delta=0	0.01	0.03	0.81

Table 8: Combined Direct-Indirect Method - Income-tax Evasion Prevalence Estimate and Hypothesis Tests – Restricted Sample (N=1046)

	Estimation	**Standard deviation**	**p-value**
Prevalence Estimate (percent)	14.7	6.54	NA
Test 1: H0: Beta=1	1.1	0.28	0.73
Test 2: H0: Delta=0	0.01	0.02	0.67

Using the restricted sample seems to have corrected the hypothesis violation problems present when using the whole sample. Tables 7 and 8 show estimates that are significantly different from the ones obtained with the simple list method (26.56 vs 28.5 for consumption taxes, $p<0.000$, and 14.69 vs 13.53 for income tax, $p<0.000$) Here again, prevalence estimates of income tax evasion using the combined method are significantly higher than estimates using the list only. The direct question estimates from these restricted samples are also not significantly different from the one obtained with the complete sample and presented in Table 1.

[11] Appendix B shows results if we had forgone this second precaution.
[12] Direct prevalence estimates for those subsamples did not significantly differ from those obtained using the full sample.

Links with compliance models

As originally stated, the objective of using a new experimental method to measure tax evasion is to gain a tool to validate our theoretical understanding of the phenomena. One of those models which has gained traction in discussions about compliance is the slippery-slope framework developed by Kirchler (2008) and multiple of his colleagues over the years. The main idea is that both coercion from authorities as well as trust in them foster a climate of compliance, either forced or voluntary. These two psychological postures can feed each other as taxpayers that believe tax authorities to be effective at catching and punishing tax evaders are likely to be more trusting of such authorities as well. Our survey included a set of attitudinal questions aimed at testing the adequacy of this model with our new data on evasion. We ran our models again but using these psychological traits as the main regressors of interest[13].

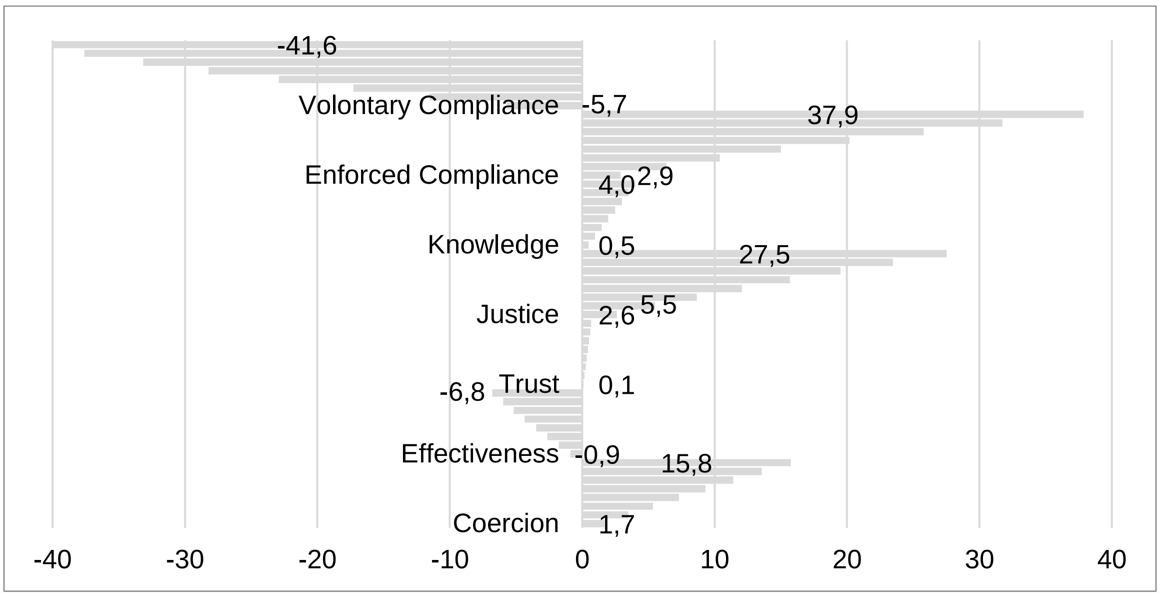

Figure 4: Psychological drivers of consumption tax evasion, average marginal effects on propensity to evade

[13] The binomial logistic regression model includes control for gender, being born in Canada, being self-employed, age and income, which are not shown in the graphs. All effects illustrated are significant at least at the 90 percent level.

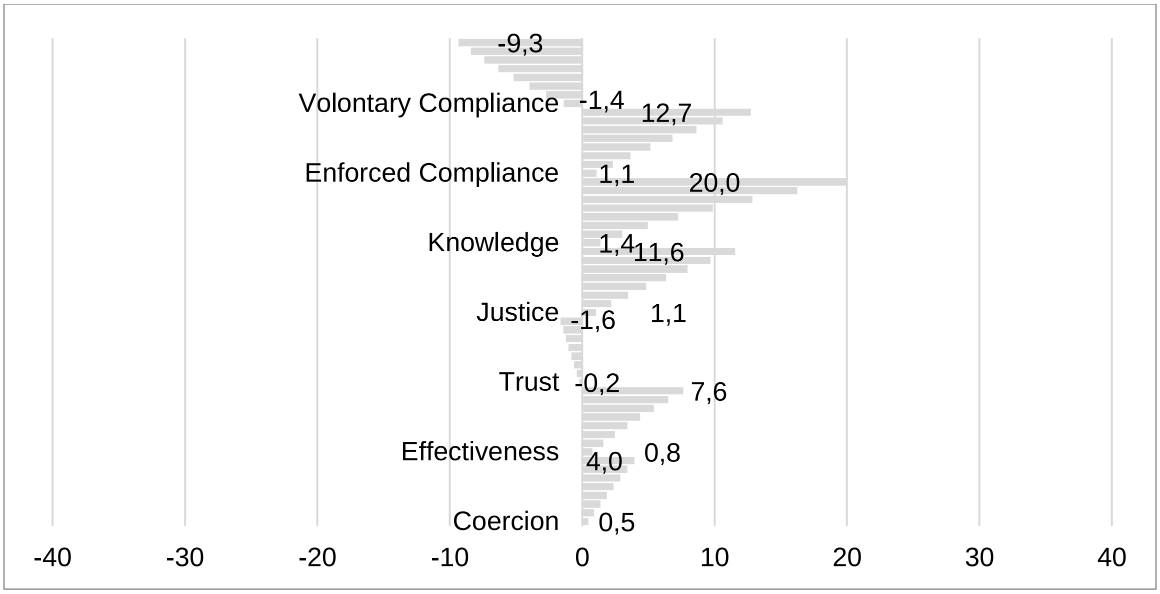

Figure 5: Psychological drivers of income tax evasion, average marginal effects on propensity to evade

We tested a version of the slippery slope model similar to the one explored by Kastlunger and her colleagues (2013)[14]. As figures 4 and 5 show, both types of compliance attitudes do not lead to compliance, with enforced compliance perception being associated with higher odds of evasion, contrary to one of the original intuitions of the model but in accordance with Kastlunger's results. All associations we find run in the same direction as those observed by these authors, even though we replaced their measure of intent or tolerance for evasion with an indirect measure of the actual act of tax evasion. We also added tax knowledge to the framework, a component of the slippery slope framework often mentioned but rarely implemented in empirical tests of it. We used a series of quiz questions about income tax as our measure of tax knowledge and, interestingly, it seems to be a driver a much stronger driver of income rather than consumption tax evasion.

Discussion

Our estimates of consumption-tax evasion are noticeably higher than the one obtained by Fergusson, Molina and Riano (2019)[15]. They find an average prevalence of 19.3% while our numbers reach 28.3% and even in the high thirties in alternative specifications. This result is somewhat surprising as Colombia has a much larger informal sector than Canada, which should produce the opposite result (Slemrod and Yitzhaki, 2002). The fact that Fergusson and colleagues asked the question in general terms while we asked specifically about evasion in the last

[14] Our results for "effectiveness" would be labelled "legitimate power" in their specification though the underlying instrument is almost the same.
[15] Castañeda, Doyle and Schwartz (2020) also used a list experiment to study tax evasion but their sensitive element is not actual evasion but intention to evade, which makes it hard to compare to both our and Fergusson, Molina and Riano's results.

year would also create the expectation of higher estimates in their study compared to ours.

As far as income tax goes, our 13.5% estimate is fairly close to Kleven and colleagues' (2011) estimates of 10.7% for Danish income taxpayers. As theory predicts, they find a much higher share of self-employed taxpayers underreport their income. Our analysis supports those findings as far as income tax goes. Sadly, the Canada Revenue Agency does not run a random audit program (CRA, 2016, CRA 2017) so there is no direct Canadian benchmark for our estimates. The CRA does run a survey program to measure tax compliance attitudes. It shows that 5% of surveyed Canadians would be very likely or likely to underreport their income for income tax purposes and that 34% of them would be likely or very likely to pay for a purchase in cash to avoid paying sales taxes (CRA 2019). These numbers are comparable in magnitude with our estimates even if they measure intentions rather than past behaviour.

Torgler (2016) suggests within-country variations in tax compliance might be useful to the study of tax morale. As far as potential provincial drivers of evasion go, British-Colombia and Quebec have the largest underground economies, Alberta has the smallest and the other provinces or regions sit in between (Statistics Canada, 2016). On top of that distinction, Quebec is the only province where taxpayers have to file two income tax declarations. The provinces also differ in the way their consumption taxes are assessed: Atlantic provinces and Ontario have provincial taxes that are harmonized with the federal value-added tax, British-Columbia, Manitoba, Saskatchewan and Quebec retain distinct provincial taxes and Albertans only pay the federal tax because they don't levy a sales' tax of their own. Links between those predictors of evasion and the results presented in Table 3 are hard to establish.

Contrary to Fergusson and colleagues who present result diverging from most research, we find associations between gender and evasion, both for consumption taxes and income tax. Women tend to confess to evading less than men in our sample, as other authors have found (Coricelli, Joffily, Montmarquette and Villeval, 2010, Slemrod, 2007). We find that age is associated with significant differences in evasion propensities but these don't follow a clear pattern. The effect of income differs by type of evasion: richer respondents have higher odds of evading income taxes and lower odds of evading consumption taxes. Fergusson and colleagues observed the same relationship, their higher-income respondents being less likely to evade consumption taxes. They also observed that education is associated with less evasion, which is the opposite of what we find for consumption taxes but cannot clearly assess for income-tax evasion.

We were the first to combine a tax evasion list experiment with questions about tax filing and motivation for evasion. Our results show that respondents that filed their taxes themselves by hand were the most likely to evade consumption taxes and those that did not file at all were the least. As far as income tax goes, having one's taxes done by a professional was associated with the lowest odds of evasion and not filing with the highest. It is difficult to make sense of the possible relationship between tax filing method and paying in cash to evade sales tax but the relationship with income tax evasion is much more obvious and the results go in the expected direction. Non-filers have, by definition, much more chance of evading taxes. The fact that respondents who had a professional file their taxes were less prone to evasion should not surprise us despite media coverage of lawyer-assisted tax evasion and avoidance. The high-income evaders and avoiders that hire firms to circumvent the tax law are probably absent in our sample. The "ordinary" tax evaders that our sample includes would most likely avoid using a tax professional since they would have to lie both to her and to the tax administration.

Our analysis also offers a few insights about the application of the list experiment method. Combining a list experiment and a direct question, as recommended by Alhquist, proved useful to support the choice of the list experiment. It gives us some assurance that, at least for income tax evasion, we were right to use an indirect questioning method. Our results, however, tend to show that having respondents see the list experiment first and the direct question second might create some bias in answers to the latter. Combining two list experiments into one survey is useful to reduce research costs but future researchers should think twice about it. If the direct estimate has value in itself, for instance if the researcher wants to study misreporting bias the way Eady did, then it might be safer to include only a single list experiment per sample.

Conclusion

Tax evasion is a very hard subject to study and multiple approaches must be employed to grasp at the issue. Between methods that try to estimate how much governments lose because of it and methods trying to understand how individuals behave in taxation experiments, we have tried to measure the phenomenon by getting survey respondents to confess their past evasion behaviour. The list experiment we used could allow us to bring external validity to results obtained in the lab about what drives the behaviour of tax evaders.

We obtained prevalence estimates for consumption taxes and income tax evasion that are comparable to results obtained using different measurement methods. One of the novelties offered by the list experiment is the possibility of coupling this information with individual information to obtain insight about predictors of tax evasion. Our results on this front show promise, but mostly in our understanding of consumption-tax evasion. Not only is this form of evasion much more common than income-tax evasion, but we also have indications that it suffers from much smaller misreporting bias. This makes correlational analysis much easier and opens the door to perhaps more direct measurement methods such as interviews or direct questioning in surveys. Income-tax evasion, on the other hand, seems to suffer from a high misreporting bias and the estimated prevalence is low, even when using an indirect questioning method. This limits the possibilities for analysis, as the low number of evaders identified makes statistical results noisier. We should, however, keep in mind that a high level of compliance for the Canadian governments' largest source of revenue is certainly a good thing.

Further research should bring together this new measurement tool with tests of theories about the drivers of tax evasion. Tax morale, inequality aversion or expected utility perceptions can all be measured in surveys and we could thus learn how good they are at predicting at least some forms of evasion. Research about the civic motivation of individuals to pay taxes should be of great interest to us. High-profile stories about tax evasion by corporations or high-wealth individuals pose the risk of undermining trust in the tax system, which could push others to evade themselves. Understanding the mechanisms that go into the decision to comply with tax obligations could help us prevent a spiral of distrust in our public institutions.

References

Ahlquist, J. S. (2018). List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators. Political Analysis, 26(1), 34–53. https://doi.org/10.1017/pan.2017.31

Allingham, M. G., & Sandmo, A. (1972). Income Tax Evasion: A Theoretical Analysis. Journal of Public Economics, 1(3–4), 323–338.

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. https://doi.org/10.1007/s10797-011-9171-2

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

Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax Compliance. Journal of Economic Literature, 36(2), 818–860. JSTOR.

Aronow, P. m., Coppock, A., Crawford, F. W., & Green, D. P. (2015). Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence. Journal of Survey Statistics and Methodology, 3(1), 43–66. https://doi.org/10.1093/jssam/smu023

Blair, G., Chou, W., & Imai, K. (2019). List Experiments with Measurement Error. Political Analysis, 27(4), 455–480. https://doi.org/10.1017/pan.2018.56

Blair, G., Coppock, A., & Moor, M. (2020). When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments. American Political Science Review, 1–19. https://doi.org/10.1017/S0003055420000374

Blair, G., & Imai, K. (2012). Statistical Analysis of List Experiments. Political Analysis, 20(1), 47–77.

Blair, G., Imai, K., & Lyall, J. (2014). Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan. American Journal of Political Science, 58(4), 1043–1063. https://doi.org/10.1111/ajps.12086

Brooks, N., & Doob, A. N. (1990). 3. Tax Evasion: Searching for a Theory of Compliant Behaviour. In M. Friedland (Ed.), Securing Compliance. University of Toronto Press. https://doi.org/10.3138/9781442679696-006

Canada Revenue Agency. (2016). Estimating and analyzing the tax gap related to the Goods and Services Tax/Harmonized Sales Tax (p. 11). Canada Revenue Agency. http://publications.gc.ca/pub?id=9.838559&sl=0

Canada Revenue Agency. (2017a). Tax Assured and Tax Gap for the Federal Personal Income Tax System (p. 55). Canada Revenue Agency. http://publications.gc.ca/pub?id=9.843144&sl=0

Canada Revenue Agency. (2017b). Evaluation study – Electronic services – Individual compliance behaviour – Tax return filing. Canada Revenue Agency.

Canada Revenue Agency. (2019). Taxpayer Attitudinal Segmentation Research 2018/2019 (p. 44).

Castañeda, N., Doyle, D., & Schwartz, C. (2020). Opting Out of the Social Contract: Tax Morale and Evasion. Comparative Political Studies, 53(7), 1175–1219. https://doi.org/10.1177/0010414019879956

Chou, W., Imai, K., & Rosenfeld, B. (2017). Sensitive Survey Questions with Auxiliary Information. Sociological Methods & Research, 004912411772971. https://doi.org/10.1177/0049124117729711

Clotfelter, C. T. (1983). Tax Evasion and Tax Rates: An Analysis of Individual Returns. The Review of Economics and Statistics, 65(3), 363–373. JSTOR. https://doi.org/10.2307/1924181

Coricelli, G., Joffily, M., Montmarquette, C., & Villeval, M. C. (2010). Cheating, emotions, and rationality: An experiment on tax evasion. Experimental Economics, 13(2), 226–247. https://doi.org/10.1007/s10683-010-9237-5

Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). Sociological Methods & Research, 40(1), 169–193. https://doi.org/10.1177/0049124110390768

Droitcour, J., Caspar, R., Hubbard, M., Parsley, T., Visschler, W., & Ezzati, T. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), Measurement Errors in Surveys (pp. 185–210). John Wiley & Sons. https://www.rti.org/publication/item-count-technique-method-indirect-questioning

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), 203–232. https://doi.org/10.1257/pol.20150083

Eady, G. (2017). The Statistical Analysis of Misreporting on Sensitive Survey Questions. Political Analysis, 25(02), 241–259. https://doi.org/10.1017/pan.2017.8

Fergusson, L., Molina, C., & Riaño, J. F. (2019). Consumers as VAT “Evaders”: Incidence, Social Bias, and Correlates in Colombia. Economía, 19(2), 21–67. https://doi.org/10.1353/eco.2019.0001

Fortin, B., Lacroix, G., & Pinard, D. (2010). Evaluation of the Underground Economy in Quebec: A Microeconomic Approach. International Economic Journal, 24(4), 463–479. https://doi.org/10.1080/10168737.2010.525983

Fortin, B., Lacroix, G., & Villeval, M.-C. (2007). Tax evasion and social interactions. Journal of Public Economics, 91(11–12), 2089–2112. https://doi.org/10.1016/j.jpubeco.2007.03.005

Gemmell, N., & Hasseldine, J. (2012). The Tax Gap: A Methodological Review. Advances in Taxation, 20(1), 203–231.

Glynn, A. N. (2013). What Can We Learn with Statistical Truth Serum? Public Opinion Quarterly, 77(S1), 159–172. https://doi.org/10.1093/poq/nfs070

Hashimzade, N., Myles, G. D., & Tran-Nam, B. (2013). Applications of behavioural economics to tax evasion: Behavioural explanations of tax evasion. Journal of Economic Surveys, 27(5), 941–977. https://doi.org/10.1111/j.1467-6419.2012.00733.x

Heerwig, J. A., & McCabe, B. J. (2009). Education and Social Desirability Bias: The Case of a Black Presidential Candidate. Social Science Quarterly, 90(3), 674–686. https://doi.org/10.1111/j.1540-6237.2009.00637.x

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. Public Opinion Quarterly, 74(1), 37–67. https://doi.org/10.1093/poq/nfp065

Imai, K. (2011). Multivariate Regression Analysis for the Item Count Technique. Journal of the American Statistical Association, 106(494), 407–416. https://doi.org/10.1198/jasa.2011.ap10415

Kastlunger, B., Lozza, E., Kirchler, E., & Schabmann, A. (2013). Powerful authorities and trusting citizens: The Slippery Slope Framework and tax compliance in Italy. Journal of Economic Psychology, 34, 36–45. https://doi.org/10.1016/j.joep.2012.11.007

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

Kleven, H. J., Knudsen, M. B., Kreiner, C., Pedersen, S., & Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. Econometrica, 79(3), 651–692.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial Attitudes and the “New South.” The Journal of Politics, 59(2), 323–349.

Lax, J. R., Phillips, J. H., & Stollwerk, A. F. (2016). Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment. Public Opinion Quarterly, 80(2), 510–533. https://doi.org/10.1093/poq/nfv056

Luks, S., & Monson, J. Q. (2010). Pushing the Limits of List Experiments [Working Paper]. Department of Political Science, Brigham Young University.

Luttmer, E. F. P., & Singhal, M. (2014). Tax Morale. Journal of Economic Perspectives, 28(4), 149–168. https://doi.org/10.1257/jep.28.4.149

Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions: Validation of Survey Methodologies for Sensitive Questions. American Journal of Political Science, 60(3), 783–802. https://doi.org/10.1111/ajps.12205

Slemrod, J. (2007). Cheating Ourselves: The Economics of Tax Evasion. Journal of Economic Perspectives, 21(1), 25–48. https://doi.org/10.1257/jep.21.1.25

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. International Tax and Public Finance, 19(1), 25–53. https://doi.org/10.1007/s10797-011-9181-0

Slemrod, J., & Yitzhaki, S. (2002). Tax Avoidance, Evasion, and Administration. In A. J. Auerbach & M. Feldstein (Eds.), Handbook of Public Economics (Vol. 3, pp. 1423–1470). Elsevier. https://doi.org/10.1016/S1573-4420(02)80026-X

Statistics Canada. (2018). The underground economy in Canada, 2016. The Daily, 11, 4.

The American Association for Public Opinion Research. (2016). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (9th ed.). AAPOR.

Torgler, B. (2016). Tax Compliance and Data: What Is Available and What Is Needed: Tax Compliance and Data. Australian Economic Review, 49(3), 352–364. https://doi.org/10.1111/1467-8462.12158

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. Psychological Bulletin, 133(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

U.S. Internal Revenue Service. (1996). Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992 (Publication 1415 (Rev. 4-96)).

Appendix A: Descriptive Table for each Experimental Path

Table 9: Distribution of respondents by characteristics and experimental path

	Full sample	Path 1	Path 2	Path 3	Path 4	Path 5	Path 6	Path 7	Path 8
Men	53.3	54.2	52.5	53.9	52	57.4	52.7	52.7	51
Women	46.7	45.8	47.5	46.1	48	42.6	47.3	47.3	49
Full-time employee	44.5	46.4	44.3	44.6	41.9	45.9	41.4	47	44.9
Part-time employee	9.3	8.9	11	10.4	6.4	9.2	11.9	8.1	8.5
Self-employed	6.3	6.4	5.8	6.4	6.7	8.4	4.5	5.5	6.7
Student	7	8.1	7.2	4.3	8.1	6.2	6.5	6.3	9
At home	3.9	3.6	3.2	4.3	5.9	3.4	4.2	4.9	1.7
Unemployed	5.7	5.3	5.2	6.1	5.6	5.3	8.2	6.3	3.5
Retired	22.6	20.4	23.2	23.2	24.9	20.4	22.7	21.6	24.8
No answer	0.6	0.8	0	0.6	0.6	1.1	0.6	0.3	0.9
Canadian Native	83.1	82.4	85.2	83.8	82.1	82.1	83.9	83.9	81.6
Non-native to Canada	16.9	17.6	14.8	16.2	17.9	17.9	16.1	16.1	18.4
18 to 24 years old	10.1	9.2	11.3	9.3	9.5	11.2	10.2	9.8	10.5
25 to 34 years old	16.9	15.9	15.7	15.4	19	15.7	15.9	19.6	18.4
35 to 44 years old	16.9	19.6	19.1	14.8	14.8	20.2	15	15	16.6
45 to 54 years old	20.2	18.7	19.7	23.2	19.6	21.6	20.7	17.6	20.4
55 to 64 years old	17.1	18.4	13.6	17.4	17.3	14.8	21.2	21.3	12.8
65 to 74 years old	13.8	13.7	16.8	15.1	13.4	10.4	12.5	12.4	16.6
75 or older	4.9	4.5	3.8	4.9	6.4	6.2	4.5	4.3	4.7
Family income below $20,000	9.1	9.5	9.3	6.4	8.7	9	11.3	7.8	10.5
20 to $40,000	17	17	16.8	18.3	17.9	17.1	17.6	16.4	15.2
40 to $60,000	18.4	15.9	19.1	20	17.3	18.5	17.8	18.7	19.8
60 to $80,000	16.1	14.8	17.4	18	16.8	14.6	12.5	16.4	19
80 to $100,000	14.3	14.2	13.9	14.5	15.6	13.4	14.4	13.8	14.3
$100,000 and above	25.1	28.5	23.5	22.9	23.7	27.5	26.3	26.8	21.3
High School Diploma or less	25.5	24.3	25.2	25.2	24	25.5	27.2	28	24.5
College Diploma	30.8	26.5	34.8	29.9	30.2	30.8	33.1	28.2	32.7
University Diploma	43.8	49.2	40	44.9	45.8	43.7	39.7	43.8	42.9

Note: p-value from χ 2 tests: Gender—0.81, Employment—0.51, Non-native to Canada—0.92, Age—0.33, Income—0.92, Education—0.52

Appendix B: Combined direct-indirect estimates using an alternative restricted sample

The sample is restricted to respondents who either were always in control groups (Paths 4 and 8) or were always in test groups (Paths 2 and 6). It pools the samples used to produce Tables 7 and 8.

Table 10: Combined Direct-Indirect Method - Consumption-tax Evasion Prevalence Estimate and Hypothesis Tests —Alternative Restricted Sample (N=1399)

	Estimation	Standard deviation	p-value
Prevalence Estimate (percent)	31.96	4.98	NA
Test 1: H0: Beta=1	0.83	0.1	0.08
Test 2: H0: Delta=0	0.02	0.02	0.37

Table 11: Combined Direct-Indirect Method – Income tax Evasion Prevalence Estimate and Hypothesis Tests —Alternative Restricted Sample (N=1399)

	Estimation	Standard deviation	p-value
Prevalence Estimate (percent)	14.98	5.15	NA
Test 1: H0: Beta=1	0.8	0.24	0.41
Test 2: H0: Delta=0	0	0.01	0.98

The combined estimator for consumption tax evasion does not clearly pass the first hypothesis under this alternative sample selection. This supports our choice of excluding respondents who saw all control or test lists but saw the income tax list first. Results for income tax evasion are very similar to those presented in Table 8, they still differ significantly from the list-only estimate ($p<0.00$) and they pass both tests.