Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

You've Got Mail: A Randomized Field Experiment on Tax Evasion

Kristina M. Bott, Alexander W. Cappelen, Erik Ø. Sørensen, Bertil Tungodden

To cite this article:
Kristina M. Bott, Alexander W. Cappelen, Erik Ø. Sørensen, Bertil Tungodden (2019) You've Got Mail: A Randomized Field Experiment on Tax Evasion. Management Science

Published online in Articles in Advance 22 Oct 2019

. https://doi.org/10.1287/mnsc.2019.3390

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

You've Got Mail: A Randomized Field Experiment on Tax Evasion

Kristina M. Bott,[a] Alexander W. Cappelen,[a] Erik Ø. Sørensen,[a] Bertil Tungodden[a]

[a] Centre for Experimental Research on Fairness, Inequality and Rationality—The Choice Lab, Norwegian School of Economics, 5045 Bergen, Norway

Contact: kristina.bott@nhh.no, https://orcid.org/0000-0002-2491-8778 (KMB); alexander.cappelen@nhh.no, https://orcid.org/0000-0002-3489-7085 (AWC); erik.sorensen@nhh.no, https://orcid.org/0000-0002-7155-4188 (EØS); bertil.tungodden@nhh.no, https://orcid.org/0000-0002-4182-8491 (BT)

Received: July 15, 2017
Revised: January 19, 2019
Accepted: April 15, 2019
Published Online in Articles in Advance: October 22, 2019

https://doi.org/10.1287/mnsc.2019.3390

Abstract. We report from a large-scale randomized field experiment conducted on a unique sample of more than 15,000 taxpayers in Norway who were likely to have misreported their foreign income. By randomly manipulating a letter from the tax authorities, we cleanly identify that moral suasion and the perceived detection probability play a crucial role in shaping taxpayer behavior. The moral letter mainly works on the intensive margin, while the detection letter has a strong effect on the extensive margin. We further show that only the detection letter has long-term effects on tax compliance.

History: Accepted by Yan Chen, behavioral economics.

Funding: This work was supported by the Research Council of Norway through its Centres of Excellence Scheme [Centre for Experimental Research on Fairness, Inequality, and Rationality (FAIR) Project 26 26 75; Grant 236995] and the New Opportunities for Research Funding Agency Cooperation in Europe (NORFACE) Welfare State Future Program [file number 462-14-030] and administered by FAIR–The Choice Lab, NHH Norwegian School of Economics.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3390.

Keywords: taxation • tax evasion • tax morale • field experiment

1. Introduction

A key challenge in all modern societies is to limit tax evasion, which causes large losses in government revenues and creates significant unfairness in society. It has, for example, been argued that the loss of government revenue amounts to 500 billion USD annually in the United States, corresponding to the size of the government deficit, and 11 billion euros in Greece, corresponding to 30% of the government deficit (Cebula and Feige 2012, Artavanis et al. 2016, Cobham and Jansky 2018). Tax evasion is particularly difficult to handle when the tax administration has to rely on self-reported data, since taxpayers have an economic incentive to underreport income (Allingham and Sandmo 1972, Slemrod and Yitzhaki 2002, Sandmo 2005). The classical approach to increasing tax compliance has therefore been to reduce the economic incentives for tax evasion by increasing the detection probability and penalties. In recent years, however, there has been a growing interest among policymakers and academics in understanding the extent to which moral motivation or, more broadly, "tax morale," can play a role in increasing tax compliance in society, since it has the potential to ensure tax compliance even in situations where the capacity to control and audit taxpayers is limited (Slemrod 2007, Luttmer and Singhal 2014).

Tax morale is an umbrella term that captures nonpecuniary motivations for tax compliance (Luttmer and Singhal 2014), including fairness concerns, social norms, and reciprocal motives (Cullis and Lewis 1997, Blumenthal et al. 2001, Torgler 2002, Wenzel 2004, Feld and Frey 2007, Kirchler 2007, Torgler 2007, Kirchler et al. 2008). Many studies have pointed to tax morale as an explanation for why the level of tax compliance is often higher than what we would expect if people were motivated only by the fear of detection and penalties (Alm et al. 1992, Andreoni et al. 1998, Bobek and Hatfield 2003), and several studies suggest that cross-country differences in tax compliance may partly reflect differences in tax morale across countries (Alm et al. 1995, Alm and Torgler 2006). Recent work has also provided neural evidence underscoring the importance of moral sentiments in the tax compliance context (Dulleck et al. 2016).

To study the drivers of tax compliance, and in particular the role of moral motivation, we conducted a large-scale field experiment together with the Norwegian Tax Administration (NTA) on a unique sample of more than 15,000 taxpayers. The sample consisted of taxpayers who were likely to have misreported their foreign income for the previous tax year, but who were not aware that the Norwegian tax authorities

had information about this misreporting. Information about foreign income is not included in the prepopulated tax return in Norway and the taxpayers therefore have to self-report this information. Historically, it has been difficult for the tax authorities to verify the self-reported information because they have not had access to third-party reports from foreign countries, but over the last years this has changed due to international collaboration among tax authorities.

The field intervention consisted of an information letter sent by the tax administration shortly before the taxpayers were due to submit their tax return for the previous year, where we randomly assigned taxpayers to receive different versions of a base letter or to a control group that did not receive any letter. The base letter contained information about why and how to report foreign income, and the effect of this letter sheds light on whether the underreporting was driven by a lack of information about tax-reporting procedures. The main aim of this study, however, was to identify the causal effects of introducing moral suasion and increasing the perceived detection probability of tax evasion, and we thus manipulated the base letter along each of these two dimensions in additional treatments. We study two versions of moral suasion—a fairness argument and a societal benefits argument—for correctly reporting foreign income. We assume that the main role of these letters is to make the moral argument for tax compliance salient, in the same way as Chetty et al. (2009) found that posting sales tax on price tags served to make the post-tax price more salient to consumers. To investigate the importance of the detection probability for tax evasion, we added information to the base letter that we believed would make the taxpayers increase their subjective probability of being audited.

We study the effect of our treatment manipulations on the self-reported foreign income both in the following tax return (the follow-up year) and one year later (long term). We also investigate whether the effects are largely on the extensive margin (how many taxpayers self-report any foreign income) or on the intensive margin (how much foreign income is self-reported by taxpayers who would have self-reported some foreign income in the absence of the treatment), and we provide evidence on the underlying mechanisms of the treatment effects.

Our main result is that moral suasion has a large and significant effect on self-reported foreign income. For the follow-up year, the average self-reported foreign income by the taxpayers who received one of the moral letters was almost double the amount self-reported by those who received the base letter. We also find a large effect of the detection letter, but the moral letters and the detection letter affect different margins of the taxpayer behavior: the detection letter has a large effect on the extensive margin, whereas the moral letters only have a large effect on the intensive margin. Further, we show that the base letter itself has some effect on self-reported foreign income, but, overall, our study suggests that the underreporting is not primarily driven by a lack of knowledge about how to report foreign income.

Our findings are robust across different subgroups (age, gender, Norwegian citizenship, socioeconomic status). For all subgroups, we observe that the moral letters and the detection letter increase the level of self-reported foreign income for the follow-up year. Finally, we study the long-term effects of the intervention, where the main insight is that the detection letter has a large effect on the extensive margin even one year after the taxpayer received the letter, whereas there are no statistically significant long-term effects of the moral letters. These long-term findings suggest that the moral letters mainly worked through making the moral arguments salient when the taxpayer received the letter, whereas the detection letter caused the taxpayers to permanently update their beliefs about the detection probability.

Our paper contributes to the growing literature using field interventions to study tax compliance (Coleman 1996, Blumenthal et al. 2001, Slemrod et al. 2001, Torgler 2004, Coleman 2007, Kleven et al. 2011, Ariel 2012, Del Carpio 2013, Fellner et al. 2013, Hallsworth 2014, Castro and Scartascini 2015, Pomeranz 2015, Dwenger et al. 2016, Hallsworth et al. 2017). Evidence from the field experimental literature on the drivers of tax evasion has been mixed and, in particular, most of the studies have not been able to document that moral suasion may play an important role in reducing tax evasion. The pioneering paper by Blumenthal et al. (2001) uses a randomized field intervention in Minnesota to investigate the effect of moral appeals. In different letters sent to taxpayers from the tax authorities, they include a message highlighting the social benefits of services financed by tax revenues and a message indicating that the majority of citizens comply with tax laws. They find no effect of these letters on tax compliance. The paper by Castro and Scartascini (2015) reports from a large field experiment in Argentina on payment of property taxes. They find no average effects on tax compliance of two messages aimed at affecting beliefs about other taxpayers' behavior and beliefs about the use of resources by the government. Similarly, Dwenger et al. (2016), in the context of German church taxes, and Fellner et al. (2013), in the context of TV licenses fees in Austria, find no effect of appealing to morals or providing information about other's compliance. However, there are also interesting studies of tax compliance in the laboratory, which provide evidence of the potential role of moral suasion and taxpayer agency in reducing tax evasion (Dal Bó and Dal Bó 2009, Lamberton et al. 2018) and of people overestimating the probability of detection (Alm et al. 1992).

We believe our study has several strengths that may contribute to explaining why we can cleanly identify strong effects of both moral suasion and information that should increase the perceived detection probability. First, we consider a sample and a situation where there is no third-party reporting and all taxpayers have an opportunity to evade taxes, whereas some previous studies have suffered from a significant part of the sample being restricted by third-party reporting (Kleven et al. 2011). Second, we carefully timed the distribution of the letters such that the taxpayers received them close to the deadline for submitting their tax returns, whereas some previous studies had a significant lag between the field intervention and the moment of decision making (Blumenthal et al. 2001, Slemrod et al. 2001). Third, our experimental design allows for a clean test of whether moral suasion or information that should increase the perceived detection probability drive the change in taxpayers' behavior. We compare the effect of the moral letters and the detection letter to the effect of a base letter that only differs in the relevant dimensions, and we provide additional survey evidence showing that the letters worked as intended. Overall, we therefore believe that our study provides novel, clean, and robust evidence of how both moral suasion and information about detection probability may contribute to reducing tax evasion. Our study also contributes to the literature by showing that moral suasion and detection probability affect different margins of taxpayer behavior, and by providing, to our knowledge, the first set of results on the long-term effects of a field intervention aiming to reduce tax evasion.

Taken together, our results shed light on the fundamental question of why people pay taxes. In particular, the results contribute to a better understanding of how tax morale affects tax compliance, and suggest that it is possible to achieve a significant increase in tax compliance by making moral appeals. However, our long-term findings show that it is not trivial to strengthen tax morale in society. Moral appeals may only have short-term effects, and more comprehensive measures are therefore most likely needed to establish and sustain a tax morale that can foster compliance.

The structure of the paper is as follows: Section 2 presents the setting for the field experiment and the sample, Section 3 provides details of the experimental design. In Section 4, we provide a simple theoretical framework to guide our analysis, and we outline the empirical strategy in Section 5. Section 6 discusses the results, and Section 7 provides some concluding remarks.

2. Background and Sample

In this section, we provide an overview of how taxes are reported in Norway. We then describe the sample of taxpayers in our study.

2.1. Tax Reporting in Norway

Every year in April, the NTA sends a prepopulated tax return for the previous fiscal year to all Norwegian tax residents. The prepopulated tax return constitutes a preliminary tax statement and the taxpayer is required to add any missing information and correct potential mistakes before the end of April. Taxpayers who believe the information in the prepopulated tax return to be correct and complete need not take any action.

When filing their taxes, taxpayers are reminded to declare all income, both domestic and foreign, earned in the previous fiscal year. The domestic income is typically for the most part included in the prepopulated tax return, based on third-party reporting in Norway, but information about foreign income must be self-reported by the tax subjects. Historically, it has been difficult for NTA to audit whether taxpayers correctly report foreign income because there has been limited exchange of information across national tax jurisdictions. Over the last few years, however, tax administrations in a number of countries have increasingly provided information about the income and wealth that tax residents of other countries earn or hold in their countries. As part of this development, NTA has in recent years received reports from other tax administrations about Norwegian tax residents' income and wealth in the respective countries. These reports are referred to as automatic country reports from abroad (ACA, in Norwegian, *Automatiske Kontrolloppgaver Utland*). The exchange of such reports is a result of bilateral negotiations between national tax authorities, and not all countries exchange this type of information with NTA.

In the period we studied, NTA received these reports with delay and thus could not include information about foreign income in the prepopulated tax return. However, NTA could compare the self-reported foreign income in the domestic tax returns with the foreign income recorded in the ACA-report at a later date, and this comparison forms the basis for our study.

2.2. The Sample

For fiscal year 2011, which is the baseline year of this study, NTA received ACA-reports for around 40,000 Norwegian tax residents. NTA estimated that 17,899 of these had self-reported between 2,000 Norwegian Krone (NOK) and 200,000 NOK (equivalent to approximately 350–35,000 USD in 2011) less in foreign income than stated in the ACA-reports for the income year 2011, and this group was the point of departure of the present study. The taxpayers were not aware that NTA had received reports about their foreign income, and the tax authorities did not act on this information until after the taxpayers had

Table 1. Background Characteristics

	Samples		With ACA-report	
Characteristic	With ACA-report	General population	Incorrect	Correct
Share Norwegian citizen	0.522	0.836	0.550	0.503
Share citizen of other Nordic country	0.456	0.039	0.431	0.474
Share female	0.445	0.502	0.455	0.437
Mean age	53.4	49.8	58.4	49.9
Share older than 60 years old	0.429	0.289	0.566	0.332
Share self-employed	0.117	0.084	0.095	0.133
Share high SES	0.487	0.353	0.452	0.512
Observations	37,897	215,956	15,708	22,189

Notes. The table reports background characteristics at baseline. "With ACA-report" refers to the individuals with an ACA-report. "General population" refers to a random sample (5%) of the population in the Norwegian tax records that do not have an ACA-report. "Incorrect" denotes the set of individuals with an ACA-report who have self-reported foreign income incorrectly, and "Correct" denotes the set of individuals with an ACA-report who have self-reported foreign income correctly. "High SES" is defined as the taxpayer being in the upper 25% of the income and wealth distribution at baseline. SES, socioeconomic status.

self-reported foreign income for 2012, which we refer to as the follow-up year in this study. A small subset of the group that self-reported incorrectly was randomly selected to be part of a practical policy experiment that tested the usefulness of a weblink providing further information about how to report foreign income. In addition, we had to exclude 137 individuals for whom NTA had incomplete ACA-reports. The remaining 15,708 individuals constitute the sample used in our analysis.

The ACA-reports suggest voluntary compliance from about 55% of the taxpayers in a situation where they have discretion to misreport. This level of voluntary compliance is in line with what Kleven et al. (2011) and Engström and Holmlund (2009) find for self-employed individuals in Denmark and Sweden, respectively, and Castro and Scartascini (2015) find for property owners in Argentina (with respect to property tax), which are groups that have substantial discretion over paying taxes.

Table 1 provides a comparison of the general population and the tax subjects with an ACA-report on foreign income at baseline (the two left columns) and a comparison of those who self-reported foreign income correctly and those who self-reported incorrectly (the two right columns), where tax subjects are classified as self-reporting correctly if they have misreported less than 2,000 NOK for the baseline year. From the left part of Table 1, we observe that compared with the general population, the tax subjects with foreign income are more likely to be non-Norwegian citizens, a large share of whom are from other Nordic countries. We also observe that the individuals with foreign income are slightly more likely to be male and self-employed and are on average a few years older than the general population. From the right part of Table 1, we observe that those who self-reported correctly and those who self-reported incorrectly are very similar with regard to background characteristics, with the exception of those who self-reported incorrectly on average being older than those who self-reported correctly.

Table 2 provides descriptive statistics on income, wealth, and misreporting at baseline. From panels A and B, we observe that the tax subjects with ACA-reports have more income and wealth than the general population. Those who self-reported incorrectly have lower income than those who self-reported correctly, whereas the two groups have the same level of wealth. Those who self-reported incorrectly have the same level of income as the general population in Norway, but substantially more wealth.

From panels C and D, we observe that average foreign income in the ACA-reports is 44,902 NOK and the mean amount misreported is 8,866 NOK. Those who self-reported incorrectly have substantially more foreign income in the ACA-reports than those who self-reported correctly (56,280 NOK versus 36,852 NOK) and self-reported only 51% of it to the tax authorities. Those who self-reported correctly reported 5,049 NOK more than what is stated in the reports from the tax administrations in other countries, which might reflect that the ACA-reports do not capture all foreign income for the tax subjects (both because the ACA-report from each country is likely to be incomplete and because NTA does not receive information from all tax authorities across the world).

3. Experimental Design

The basic structure of the experimental design is that all individuals in our sample received a prepopulated tax return for the follow-up year in week 14 of 2013, and individuals in the treatment arms then received a letter from the Norwegian tax authorities in week

Table 2. Income, Wealth, and Misreporting (Baseline)

	Samples		With ACA-report	
	With ACA-report	General population	Incorrect	Correct
Panel A: Taxable income				
Mean	360,628	272,616	299,838	403,619
Q_{25}	119,834	110,447	97,199	147,551
Q_{50}	234,809	215,354	182,845	274,685
Q_{75}	419,110	345,076	347,295	458,413
Panel B: Taxable wealth				
Mean	1,330,938	462,820	1,530,805	1,189,590
Q_{25}	0	0	0	0
Q_{50}	43,248	63	58,366	35,277
Q_{75}	609,583	325,706	651,296	577,269
Panel C: ACA-reports of foreign income				
Mean	44,902		56,280	36,852
Q_{25}	519		7,509	105
Q_{50}	6,560		18,987	868
Q_{75}	29,073		48,670	12,284
Panel D: Estimate of misreporting				
Mean	8,866		28,533	–5,049
Q_{25}	16		4,187	3
Q_{50}	985		14,209	154
Q_{75}	13,556		36,732	948

Notes. Panels A and B refer to (taxable) income and wealth at baseline. Panel C shows total foreign income in the ACA-reports, and panel D shows misreported foreign income (calculated by the difference between the foreign income in the ACA-reports and self-reported foreign income). The term Q_x refers to the x-percentile in the relevant group. "With ACA-report" refers to individuals with an ACA-report. "General population" refers to a random sample (5%) of the population in the Norwegian tax records that do not have an ACA-report. "Incorrect" denotes the set of individuals with an ACA-report who have self-reported foreign income incorrectly, and "Correct" denotes the set of individuals with an ACA-report who have self-reported foreign income correctly.

15 about how to report foreign income in the tax return. The experiment has an intention-to-treat design, since we do not know how many of the taxpayers actually read the letter. The likelihood of reading the letter, however, should not differ across treatments, since there were no treatment differences in the design of the envelopes. At our request, NTA conducted an independent survey to test whether tax subjects in general read letters from the tax authorities. The results from this survey suggest that about 60%–70% of taxpayers read the letters they receive from NTA.

The individuals could make changes to the prepopulated tax return, including self-reporting of foreign income, until the deadline in week 18. The individuals were randomly allocated into a control group (*No letter*) or one of three treatment arms (*Base* treatment, *Moral* treatments, or *Detection* treatment). The individuals in the control group did not receive any letter from the tax authorities, whereas the individuals in the different treatment arms received different versions of the base letter. In the online appendix, we provide translated versions of all the letters (see Figures B1–B4). Our main interest is whether receiving such a letter increased the self-reported foreign income for the follow-up year. We also have long-term data that allow us to study whether the letter intervention affected the self-reported foreign income one year later.

All the letters had a link to the NTA website and the phone number of a call center in the tax administration established for the purpose of this study. To standardize the answers to the taxpayers interacting with the call center, NTA provided phone operators with a script of potential questions and answers. The phone operators were not aware that the call center was facilitating a field experiment; they only knew that the authorities had sent out different letters to different individuals. In Table 3, we provide an overview of the activity at the call center. Five percent of the individuals receiving a letter approached the call center, mainly asking questions about why they had received the letter and how to report foreign income. Significantly more individuals used the call center in the *Detection* treatment than in the Base treatment (13.1% versus 3.6%, $p < 0.001$), whereas we see no difference between the *Moral* treatments and the Base treatment (3.7% versus 3.6%, $p = 0.710$).

3.1. *Base* Treatment

Since any letter from the tax authorities may cause a change in behavior for a number of reasons (fear of detection, moral salience, or better knowledge about how to proceed with reporting), we included a treatment where the tax residents received a letter that only contained general information about how to self-report foreign income (*Base* treatment).

The letter consists of three paragraphs, the first explaining why the reader is receiving this letter. It refers to the fact that the Norwegian economy is becoming more international, with an increasing number of taxpayers having income from abroad, and states that NTA would like to inform taxpayers about how this type of income is taxed and how it should be reported. The second paragraph states that all Norwegian tax residents are liable to pay taxes to Norway on all income and assets, even on foreign income and foreign assets, unless otherwise specified in Norway's tax treaties with other countries. The final paragraph describes how to proceed after receiving the prepopulated tax return, and adds a weblink providing further information about how to file the Norwegian tax return. We had two different versions of the base letter, one using active language, thus, addressing the reader as "you," and another using passive language (Bryan et al. 2012). We do not find an economically or statistically significant

Table 3. Statistics from the Contact Logs

		Treatment group			Test of equality p-values	
Statistic	Total	*Base*	*Moral*	*Detection*	*Moral*	*Detection*
Share of letter recipients in the logs	0.050	0.036	0.037	0.131	0.710	<0.001
Communication with letter recipient:						
Question about why he/she receives letter	0.282	0.261	0.254	0.324	0.888	0.182
Negative reaction to letter	0.055	0.049	0.041	0.073	0.701	0.364
Question about rules of taxation	0.347	0.348	0.351	0.343	0.951	0.936
Arguing about legitimacy of rules	0.063	0.070	0.058	0.065	0.621	0.832
Question about how to report foreign income	0.721	0.741	0.758	0.668	0.709	0.126
Question about how foreign income is audited	0.019	0.007	0.014	0.031	0.540	0.127
Question about reporting income from before 2012	0.062	0.035	0.041	0.099	0.762	0.021
Characteristics of caller:						
Mean age	67.3	70.4	68.3	64.4	0.105	<0.001
Share female	0.544	0.643	0.558	0.475	0.089	<0.001
Share Norwegian citizen	0.611	0.664	0.605	0.589	0.223	0.138
Observations	700	143	294	263		

Notes. The table shows the share of taxpayers that contacted the call center, the share that asked each of the questions, and background characteristics. The table is based on 700 contacts with the call center in which callers identified which letter they had received. "Test of equality" reports a test of equality between the respective treatment and the *Base* treatment. The p-values are based on Pearson χ tests on binary outcomes (and t-tests for tests of mean age).

difference between these two versions of the base letter on the amount self-reported ($p = 0.775$) or on the share of individuals self-reporting a positive amount ($p = 0.884$), and thus we pool them in the analysis.

The *Base* treatment allows us to study whether lack of information about how to report foreign income is a main driver of the observed underreporting of foreign income. It is, however, important to note that the treatment difference between the *Base* treatment and the *No letter* group provides an upper bound estimate of the role of information, since the base letter may also trigger other mechanisms among taxpayers.

3.2. *Moral* and *Detection* Treatments

To identify as cleanly as possible the causal effects of introducing moral suasion and of increasing the detection probability, we manipulated the base letter along each of these two dimensions in additional treatments. The treatment manipulations only introduced minor changes in the first paragraph of the letter; the rest of the letter was identical to the base letter.

We studied two types of moral appeals. In the *Fairness* treatment, the letter introduced a fairness argument for reporting foreign income correctly by including a sentence that reminded taxpayers that most Norwegians report the income earned in Norway correctly. Specifically, the following two sentences were added to the end of the first paragraph: "The great majority report information about their income and assets in Norway correctly and completely. In order to treat all taxpayers fairly, it is therefore important that foreign income and foreign assets are reported in the same manner." The statement in this letter is correct. Most income earned in Norway is third-party reported to the tax authorities, and thus Norwegians have little discretion to evade taxes. An estimate of the share of individuals in Norway having all their income third-party reported is provided in Foss et al. (2015), who analyze a representative sample of 4,000 tax returns. They show that about 70% of Norwegians did not make adjustments to the prefilled tax return in 2010 and establish that there are very few errors in these tax returns, suggesting that these taxpayers only have third-party reported income. The numbers are in line with what is found for Denmark in Kleven et al. (2011), where 62% did not self-report any additional income (p. 665, panel B). The difference between the estimates of Foss et al. (2015) and Kleven et al. (2011) is consistent with technological progress making third-party reporting increasingly common.

In the *Societal benefits* treatment, the letter introduced a societal benefits argument for reporting foreign income correctly by including a sentence that reminded taxpayers about the benefits to society resulting from taxation: "Your tax payment contributes to the funding of publicly financed services in education, health and other important sectors of society." This sentence may trigger a reciprocity motive for tax compliance, where individuals become more willing to pay taxes because they recognize this as an exchange for benefits that the state provides (Fehr and Gächter 1998, Luttmer and Singhal 2014). In two additional treatments, we visualized the societal benefits from taxation by adding an attachment

illustrating publicly financed services in health, education, infrastructure, and research (see Figure B5 in the online appendix). In one treatment, the attachment was combined with the base letter; in another treatment it was combined with the societal benefits letter.

In the *Detection* treatment, the letter aimed to increase the perceived detection probability of the tax subject. We replaced the first sentence in the base letter, "The Norwegian economy is becoming more internationalised, and an increasing number of Norwegian taxpayers receive income and have assets abroad," with, "The tax administration has received information that you have had income and/or assets abroad in previous years." The basic idea behind this treatment manipulation was that providing information about the tax authorities' knowledge about the individual's income-generating activities abroad in previous years would make the tax subjects update their subjective beliefs about the likelihood of being audited. However, it is not obvious that the detection letter has a strong effect on tax compliance. The taxpayers may believe that the tax authorities only have information about some of their foreign income or they may not consider the letter to represent a credible threat about tax enforcement (Slemrod et al. 2001).

Table 4 provides an overview of the different treatment arms in the experiment. The experimental design allows for the following two main comparisons to study the drivers of the misreporting of foreign income:

- The role of moral motivation: The comparison between the *Moral* treatments and the *Base* treatment identifies the causal effect of moral suasion on taxpayer behavior.
- The role of the detection probability: The comparison between the *Detection* treatment and the *Base* treatment identifies the causal effect of increasing perceived detection probability on taxpayer behavior.

When interpreting these treatment comparisons, we make two assumptions. First, we assume that the moral letters only manipulate the moral dimension relative to the base letter. Second, we assume that the detection letter only manipulates the perceived detection probability relative to the base letter. We tested these assumptions in an independent survey, where, as shown in Table A1 in the online appendix, we find support for the letters working as intended. Importantly, we find no evidence of the moral letters increasing the perceived detection probability among recipients of these letters.

4. Theoretical Framework

We here provide a simple model of taxpayer behavior to guide our analysis and the interpretation of the results, building on Cappelen et al. (2007) and Sandmo (2012). We do not model the behavior of the tax authorities and thus do not consider the possibility of strategic interaction between the taxpayers and the tax authorities (Graetz et al. 1986).

Assume that the taxpayer has (only) foreign income y and self-reports the foreign income to be r. The tax on foreign income is t and the penalty on misreported income, if detected, is τ. After tax, income is $Y = y - tr$ if the misreporting is not detected and $Z = y - tr - \tau(y - r)$ if it is detected. Let us assume that the taxpayer has the following expected utility function:

$$EU(r;\cdot) = pu(Z) + (1-p)u(Y) - s(T)\beta(y-r)^2, \quad (1)$$

where p is the subjective probability of being detected, u is a von Neumann-Morgenstern utility function capturing the risk preferences over income, $s(T) \in [0,1]$ is a parameter capturing the salience of the moral cost in treatment T, and $\beta \geq 0$ is the weight attached to the morals cost of misreporting when the moral cost is fully salient ($s(T) = 1$). The model captures that people find it morally right to report their income correctly and that the moral cost is increasing in the deviation between their true income and their

Table 4. Overview of Treatments

Treatment	Description
No letter	Did not receive a letter
Base	General information letter
Fairness	Base letter + the following sentence added to the first paragraph: "The great majority report information about their income and assets in Norway correctly and completely. In order to treat all taxpayers fairly, it is therefore important that foreign income and foreign assets are reported in the same manner."
Societal benefits	Base letter + the following sentence added to the first paragraph: "Your tax payment contributes to the funding of publicly financed services in education, health and other important sectors of society." Two additional treatments included an attachment visualizing public services financed through taxes (without the base letter/in combination with the base letter).
Detection	Base letter, but the first sentence ("The Norwegian economy is becoming more internationalized and an increasing number of Norwegian taxpayers receive income and have assets abroad") is replaced by the following sentence: "The tax administration has received information that you in previous years have had income and/or assets abroad."

reported income. It follows straightforwardly that the interior solution for the taxpayer is given by

$$r = y - \frac{\Delta u^s(r,t,\tau,p)}{2s\beta}, \tag{2}$$

where $\Delta u^s(r,t,\tau,p) = t(1-p)u'(Y) - (\tau - t)pu'(Z)$. The second term in the first order condition captures the trade-off that determines the level of misreporting. The nominator represents the marginal gain of misreporting at the interior solution, whereas the denominator shows the marginal cost. In the case where the moral cost is salient ($s > 0$) and the taxpayer only cares about the moral cost ($\beta \to \infty$), the optimal choice is to self-report the foreign income correctly.

The *Moral* treatments and the *Detection* treatment aim to increase the weight attached to the moral cost of misreporting, $s\beta$, and the subjective detection probability p, respectively, and thereby to increase the self-reported foreign income r. We do not expect the moral letters to fundamentally change the strength of moral motivation β, but they may make the moral argument more salient, causing an increase in s. The *Detection* treatment provides new information to taxpayers, which should make them update their subjective beliefs about the likelihood of being detected, causing an increase in p.

The treatments may work both on the extensive and the intensive margin, but they do not necessarily work on the same margin. Let us provide an example. Assume that there are two types of taxpayers: *selfish* ($\beta = 0$) and *moral* ($\beta > 0$). Further, assume that everyone has a subjective probability of detection $p = 0$ in the *Base* treatment, which implies that the selfish taxpayers do not report any foreign income. Finally, assume that the moral taxpayers are at an interior solution and report some foreign income. In this case, it follows that the *Moral* treatments only work at the intensive margin: they increase the foreign income reported by the moral taxpayers by making the moral cost more salient, whereas the selfish individuals will not be affected by an increase in the salience parameter s. In contrast, the *Detection* treatment affects both the extensive and the intensive margin. It affects the extensive margin by increasing the subjective probability of detection and thereby making it more costly for the selfish taxpayers not to report any foreign income, and it affects the intensive margin since moral taxpayers also care about the subjective probability of detection.

The main insights from this example apply to all situations where we have a separation of selfish taxpayers not reporting any foreign income and moral taxpayers reporting some foreign income. In all such cases, only the *Detection* treatment works on the extensive margin, whereas the *Moral* treatments only work on the intensive margin. More generally, the treatment effects will depend on the tax parameters, the shape of the utility function, the subjective detection probability, and the salience and importance assigned to the moral cost of misreporting.

5. Data and Empirical Strategy

We here provide an overview of the data and the empirical strategy for the main analysis and the heterogeneity analysis.

5.1. Data

The analysis uses data from the administrative records of NTA. Self-reported foreign income in the 2011 tax return serves as a baseline variable, whereas the main outcome variable of interest is self-reported foreign income for 2012 (follow-up year) and 2013 (long term). Further, we use the ACA-reports received by NTA for 2011 and 2012 from 17 countries as

Table 5. Baseline Characteristics by Treatment

Treatment	n	*Positive amount self-reported for 2011*	*Amount self-reported for 2011*	*Amount in ACA-reports for 2011*	*Amount in ACA-reports for 2012*	*Share female*	*Share above 60 years old*	*Share Norwegian citizen*	*High socioeconomic status*
No letter	1,968	0.393	30,287	83,995	43,303	0.460	0.544	0.553	0.441
		(0.011)	(4,356)	(33,390)	(6,761)	(0.011)	(0.011)	(0.011)	(0.011)
Base	3,947	0.402	27,427	55,211	44,104	0.450	0.551	0.553	0.454
		(0.008)	(3,278)	(3,935)	(5,632)	(0.006)	(0.008)	0.008)	(0.008)
Moral	7,821	0.405	28,040	51,971	38,097	0.459	0.561	0.548	0.454
		(0.006)	(2,784)	(2,070)	(2,637)	(0.06)	(0.006)	(0.005)	(0.006)
Detection	1,972	0.417	24,689	47,855	58,313	0.446	0.542	0.548	0.452
		(0.011)	(3,436)	(2,777)	(20,441)	(0.011)	(0.011)	(0.011)	(0.011)
Total	15,708	0.404	27,747	56,280	42,796	0.455	0.554	0.550	0.452
		(0.004)	(1,756)	(4,434)	(3,321)	(0.004)	(0.004)	(0.004)	(0.004)

Notes. The table reports the baseline characteristics used as controls in the main regression specifications, by treatment. Standard errors of the means are in parentheses.

background variables. According to our agreement with NTA, we are not allowed to list the countries providing ACA-reports to NTA. The administrative records also include data on age, gender, citizenship, income, and wealth.

Table 5 shows the background variables used in the analysis by treatment. We note that there are some differences across treatments in the amounts self-reported for 2011 and in the amounts in the ACA-reports for 2011 and 2012, but these differences are not statistically significant (one-way analysis of variance [ANOVA] test: amount reported 2011, $p = 0.88$; amount in ACA-reports 2011, $p = 0.12$; amount in ACA-reports 2012, $p = 0.29$.) For all other background variables, there is almost no variation across treatments.

5.2. Empirical Strategy

In the analysis, our main regression specification is

$$y_{i,t} = \alpha + \sum_{l \in L} \beta_l d_{il} + \delta y_{i,t_0} + \gamma \mathbf{x}_i + \varepsilon_i, \tag{3}$$

where $y_{i,t}$ is self-reported foreign income for individual i for year t. We let l index a treatment in the set of treatments L, where d_{il} is an indicator variable for whether individual i is in treatment l; y_{i,t_0} is the self-reported foreign income for the baseline year t_0, and $\mathbf{x}_i$ is a vector of background variables (including age, gender, Norwegian citizenship, and a measure of socioeconomic status defined by income and wealth). The estimated causal effect of treatment l relative to the reference treatment is then given by the estimated β_l coefficient.

We will start by reporting regressions where we pool all the treatment arms and define receiving a letter as the omitted category, which provides us with an estimate of the average causal effect of receiving any one of the letters from the tax authorities. We then report regressions where we estimate separately the average causal effects of the different letters relative to the *Base* treatment, before we pool the *Moral* treatments to get an estimate of the average causal effect of the *Moral* treatments. We use the same approach to study the effect on the extensive margin, where the dependent variable is an indicator variable for self-reported foreign income for year t being strictly positive. For all specifications, we report estimates for regressions both with and without the background variables. We also report the effect on self-reported foreign income conditional on it being positive, but it is important to keep in mind that this conditional variable is determined both by the extensive margin and the intensive margin.

If we are willing to make the reasonable assumption of monotone treatment response with respect to a *Base* treatment (b), we can be more specific about the effect on the intensive margin (Manski 1997). Let y_i^l denote the self-reported foreign income of individual i in treatment l, and assume a monotone treatment response where $y_i^l \geq y_i^b$ for all individuals i in treatment l. This assumption, combined with random assignment to treatment, allows us to decompose the average treatment effect as $\mathrm{ATE}^l = \pi^b \mathrm{ATE}_b^l + \pi^{\Delta(l)} \mathrm{ATE}_{\Delta(l)}^l$, where π^b is the share of taxpayers who reports a positive amount in the *Base* treatment, and ATE_b^l is the average treatment effect on this group–the intensive margin–in treatment l. The proportion $\pi^{\Delta(l)}$ represents those that did not report anything in the *Base* treatment, but that are induced to do so in treatment l, and is the effect on the extensive margin: $\pi^{\Delta(l)} = \Pr(y_i > 0 | L = l) - \Pr(y_i > 0 | L = b)$. For a treatment for which there is no effect on the extensive treatment, $\pi^{\Delta(l)} = 0$, it follows that the effect on the intensive margin is given by $\mathrm{ATE}_b^l = \mathrm{ATE}^l / \pi^b$. For a treatment with a positive extensive margin, $\pi^{\Delta(l)} > 0$, the effect on the intensive margin cannot be point estimated. However, since we assume $y_{i,t}^l \geq y_{i,t}^b$, it follows that $\mathrm{ATE}_{\Delta(l)}^l \geq 0$, and we can bound the intensive margin: $\mathrm{ATE}_b^l \in [0, \mathrm{ATE}^l / \pi^b)$. Being more specific about the intensive margin with nonzero extensive margin effects would rely on parametrically modeling responses, and, as with other selection models, identification would rely either on fully specifying the distributions of all unobservable variables or on access to a variable that predicts reporting a positive amount without directly affecting the amount itself (Staub 2014).

To study whether there are large differences in how the treatments affect the participants, we also conduct a heterogeneity analysis using the background variables age, gender, Norwegian citizenship, and socioeconomic status. In this analysis, we take the regression specification where we have pooled the *Moral* treatments as the point of departure. For each background variable, we partition the set of participants I into G and $I \setminus G$, with g_i as an indicator variable for whether individual i is a member of G. To illustrate, if the relevant background variable is age, then we partition the set of participants into two subsets, those who are below and above 60 years. The indicator variable would then take the value one if the taxpayer is above 60 years. In each case, we interact the indicator variable with the treatment indicator d_{il},

$$y_{i,t} = \alpha + \beta d_{il} + \theta g_i d_{il} + \lambda g_i + \delta y_{i,t_0} + \gamma \mathbf{x}_i + \varepsilon_i. \tag{4}$$

In this analysis, the estimation sample is the participants in the l treatment and in the *Base* treatment. With this specification, the estimated treatment effect of being in treatment l for individuals in group G is $\beta + \theta$, whereas it is β for individuals in group $I \setminus G$. The parameter θ is the estimated difference in treatment

effect between the two groups, and it provides the basis for a statistical test of whether the estimated heterogeneity is statistically significant.

6. Results

In this section, we examine how the treatments affected self-reporting for the follow-up year, before turning to the heterogeneity analysis. In the final part, we report on long-term effects of the intervention.

6.1. Main Analysis

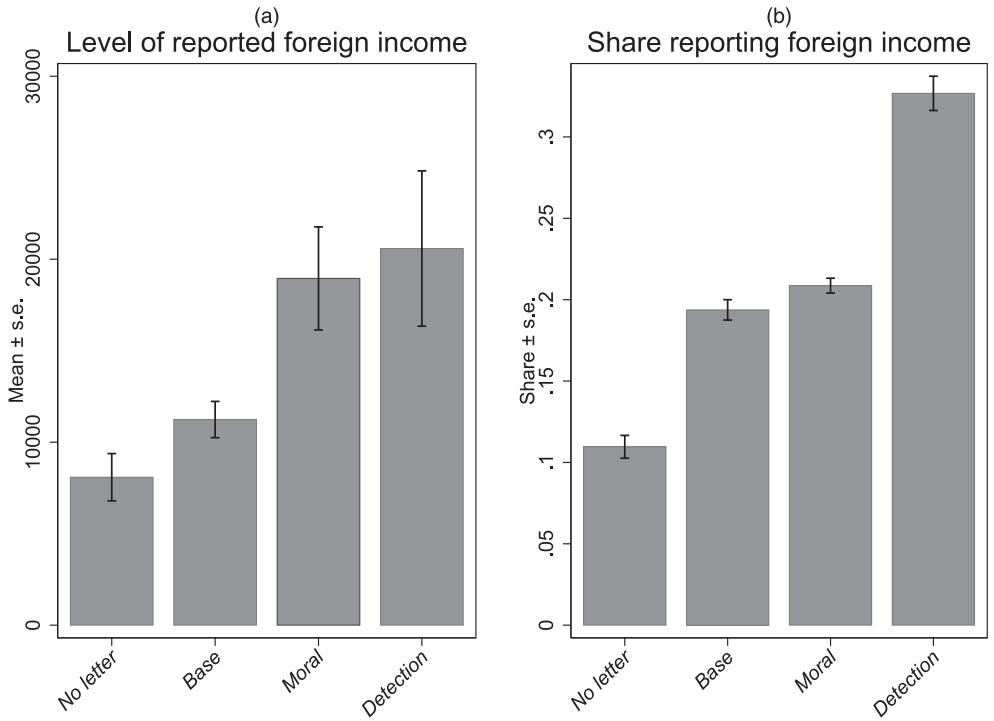
From the ACA-reports, we know that 78% of the taxpayers in our sample had foreign income in the follow-up year, on average 42,796 NOK. We now study whether the different letters caused the taxpayers to self-report a larger share of their foreign income to tax authorities.

As shown in panel (a) of Figure 1, for the follow-up year, the average self-reported foreign income of the taxpayers who received the base letter is larger than that of those who did not receive any letter. Strikingly, however, we observe that the self-reported foreign income of the taxpayers who received one of the moral letters or the detection letter is almost double the amount self-reported by those who received the base letter. In Table 6, we report the corresponding regressions based on Equation (3). Columns (1) and (2) report estimates of the effect of not receiving a letter, where all taxpayers who did receive a letter are pooled and serve as the reference category. We observe from column (1) that receiving a letter has a large and highly statistically significant effect on self-reported foreign income: it more than doubles, from 8,155 NOK in the *No letter* group to 17,030 NOK in the treatment groups combined. As shown in column (2), the finding is robust to the inclusion of a set of background variables, including self-reported foreign income for the baseline year, foreign income recorded in the ACA-report for the baseline year and the follow-up year, and personal and socioeconomic characteristics of the taxpayer ($p < 0.001$, column (2)). Finally, in panel (a) of Figure 2, we show that there are no statistically significant differences across treatments in the amount requested for deductions based on taxes paid abroad ($p = 0.551$), even though panel (b) shows that the letters cause a slight increase in the share of taxpayers requesting a deduction. Thus, we can state our first main result.

Result 1. *A letter from the tax authorities has a large and statistically significant effect on the self-reported foreign income.*

In columns (3) and (4), we report estimated treatment effects for each of the letters separately, where the base letter now serves as the reference treatment. We observe that the *No letter* treatment group reports a lower foreign income than the *Base* treatment group.

Figure 1. Self-Reported Foreign Income



Note. The figure shows, for the follow-up year, the average amount of self-reported foreign income in NOK (panel (a)) and the share of taxpayers who self-report a positive foreign income (panel (b)), by treatment.

Table 6. Treatment Effects on Self-Reported Foreign Income

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
No letter	−8,874.9***	−10,008.1***	−3,188.0*	−4,339.1	−3,188.0*	−4,339.6
	(2,184.5)	(2,767.4)	(1,643.1)	(2,734.8)	(1,643.1)	(2,735.2)
Fairness			15,158.5*	10,372.1**		
			(8,860.6)	(4,041.9)		
Societal benefits			5,180.9**	6,345.5**		
			(2,596.8)	(2,989.0)		
Detection			9,199.6**	10,351.3**	9,199.6**	10,351.6**
			(4,385.6)	(4,702.7)	(4,385.5)	(4,702.8)
Moral					7,671.1**	7,350.5***
					(3,010.7)	(2,772.9)
Foreign income baseline		0.42**		0.42**		0.42**
		(0.20)		(0.20)		(0.20)
ACA-report baseline		0.00089		0.00091		0.00091
		(0.0035)		(0.0035)		(0.0035)
ACA-report follow-up		0.010		0.010		0.010
		(0.0073)		(0.0073)		(0.0073)
Female		6,791.0*		6,782.8*		6,777.1*
		(3,640.9)		(3,634.1)		(3,633.9)
Age > 60 years		9,167.8*		9,148.3*		9,144.6*
		(5,207.1)		(5,209.7)		(5,207.6)
Norwegian citizen		2,702.2		2,733.2		2,741.5
		(3,294.8)		(3,307.6)		(3,306.1)
High SES		4,591.3		4,615.8		4,595.0
		(5,275.6)		(5,262.7)		(5,274.6)
Constant	17,029.8***	−6,706.3	11,342.9***	−12,387.4	11,342.9***	−12,380.8
	(1,752.6)	(6,673.8)	(999.6)	(7,537.0)	(999.6)	(7,535.5)
p-value from *F*-test on *Moral* treatments being equal:			0.27	0.32		
Observations	15,708	15,708	15,708	15,708	15,708	15,708
R^2	0.000	0.230	0.001	0.231	0.001	0.231

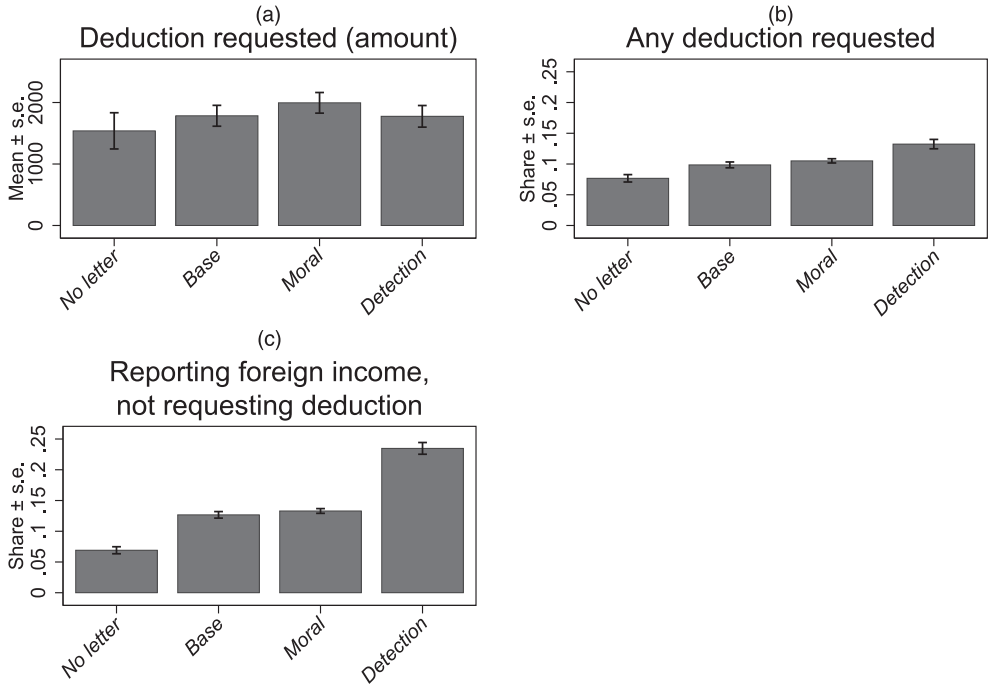
Notes. The table reports regressions based on Equation (3), where the dependent variable is the amount of foreign income self-reported for the follow-up year. In columns (1) and (2), the estimated effects are relative to the pooled sample of all treatment groups; in columns (3)–(6), the estimated effects are relative to the *Base* treatment. The indicator variables *No letter*, *Fairness*, *Societal Benefits*, and *Detection* take the value one if the taxpayer is in the respective treatment. The indicator variable *Moral* takes the value one if the taxpayer is in the *Fairness* treatment or the *Societal Benefits* treatment. The reported *p*-value from *F*-test is for the hypothesis that all *Moral* treatments have the same effect. Columns (2), (4), and (6) include the following controls: the amount of self-reported foreign income for the baseline year, the amount of foreign income recorded in the ACA-reports for 2011 and 2012 (scaled in units of one million NOK), gender, age (an indicator variable taking the value one if the taxpayer is more than 60 years old), an indicator variable taking the value one if the taxpayer is a Norwegian citizen, and an indicator variable of socioeconomic status taking the value one if the taxpayer is in the upper 25% of the income and wealth distribution in the baseline year. SES, socioeconomic status.

Robust standard errors are in parentheses: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The estimated effect of the base letter is positive, but not statistically significant ($p = 0.113$, column (4)), which suggests that the underreporting is not primarily driven by a lack of information about how to report foreign income.

The estimates for the moral letters and the detection letter identify the causal effects of adding moral suasion and increasing the detection probability. We observe that the effects are large and highly significant for all the three main treatments (*Fairness*: $p = 0.010$; *Societal benefits*: $p = 0.034$; *Detection*: $p = 0.028$, column (4)).[1] Moral suasion, in terms of a fairness or societal benefits argument, has a strong positive effect on self-reported foreign income of almost the same magnitude as information that should increase the perceived detection probability. The argument presented in the *Fairness* treatment explicitly appeals to the importance of treating all taxpayers fairly, and our results are thus in line with previous research in behavioral economics documenting that people are willing to sacrifice pecuniary gains to avoid large deviations from what they consider fair (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Andreoni and Miller 2002, Charness and Rabin 2002, Cappelen et al. 2007). However, the *Fairness* treatment also introduces a social comparison by emphasizing what the majority of

Figure 2. Requests for Deductions of Taxes Paid Abroad



Note. The figure shows, for the follow-up year, the average amount of requested deductions in Norwegian taxes in NOK (panel (a)), the share that requested deductions (panel (b)), and the share that reported positive amounts of foreign income without requesting any deductions (panel (c)), by treatment.

taxpayers do, and this social comparison could also increase compliance by appealing to a desire to conform to the social norm that tax compliance is socially considered as appropriate behavior (Cullis and Lewis 1997, Wenzel 2004, Bartke et al. 2017). In this respect, the *Fairness* treatment may affect tax compliance through both a focusing effect, by drawing taxpayers' attention to the norm, and an information effect, by making taxpayers aware of how many others follow the norm (Krupka and Weber 2009).

As shown in columns (5) and (6) of Table 6, where we pool the *Moral* treatments, moral suasion on average causes an increase in self-reported foreign income of almost 70% ($p = 0.008$, column (6)). The *Detection* treatment increases average self-reported foreign income by 80% ($p = 0.028$, column (6)).

Result 2. *Including moral suasion or information that increases the detection probability in the letter from the tax authorities has an economically and statistically highly significant effect on the amount of self-reported foreign income.*

The strong effect of the moral appeals in our study is likely to reflect the fact that we consider a sample of taxpayers who had the opportunity and willingness to evade taxes, as well as the fact that the taxpayers were exposed to the moral appeals close to the deadline for submitting their tax return. Furthermore, context-specific factors might have made the moral appeals particularly effective in our experiment. In particular, the social benefit argument and the fairness argument are likely to be more effective in a country where it is generally recognized that taxes finance important public goods and where the general level of compliance is high (Mascagni 2018). The high level of trust in Norway might also have contributed to making the moral appeals more effective than what is found in previous studies (Kirchler 2007, Kirchler et al. 2008).

In panel (b) of Figure 1, we show the effect of the letters on the extensive margin. We observe that the base letter significantly increases the share of taxpayers who report a positive foreign income for the follow-up year. The moral letter only has a small effect on the extensive margin compared with the base letter, whereas the detection letter causes a large increase in the share reporting a positive amount. Thus, consistent with our theoretical framework, the *Moral* and *Detection* treatments affect the extensive margin very differently. In Table 7, we report the corresponding regression analysis. In columns (1) and (2), we observe that receiving a letter from the tax authorities increases the share of taxpayers who report a positive foreign income from 11% in the *No letter* group to 22% in the treatment groups combined, and, as shown in column (2), the effect is robust to the inclusion of the set of background variables ($p < 0.001$, column (2)).

Table 7. Treatment Effects on Self-Reporting of Any Foreign Income

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
No letter	−0.11***	−0.11***	−0.086***	−0.083***	−0.086***	−0.083***
	(0.0079)	(0.0076)	(0.0095)	(0.0091)	(0.0095)	(0.0091)
Fairness			0.031***	0.030***		
			(0.011)	(0.011)		
Societal benefits			0.0086	0.0056		
			(0.0082)	(0.0079)		
Detection			0.13***	0.13***	0.13***	0.13***
			(0.012)	(0.012)	(0.012)	(0.012)
Moral					0.014*	0.012
					(0.0078)	(0.0075)
Positive foreign income baseline		0.16***		0.16***		0.16***
		(0.0073)		(0.0073)		(0.0073)
ACA-report baseline		0.022*		0.022*		0.022*
		(0.012)		(0.013)		(0.013)
ACA-report follow-up		0.031		0.029		0.029
		(0.023)		(0.024)		(0.023)
Female		0.052***		0.052***		0.052***
		(0.0063)		(0.0063)		(0.0063)
Age > 60 years		0.21***		0.21***		0.21***
		(0.0075)		(0.0075)		(0.0075)
Norwegian citizen		0.00033		0.00014		0.00019
		(0.0069)		(0.0068)		(0.0068)
High SES		0.037***		0.038***		0.038***
		(0.0070)		(0.0070)		(0.0070)
Constant	0.22***	−0.0035	0.20***	−0.029***	0.20***	−0.029***
	(0.0035)	(0.0065)	(0.0063)	(0.0081)	(0.0063)	(0.0081)
p-value from *F*-test on moral treatments being equal:			0.038	0.021		
Observations	15,708	15,708	15,708	15,708	15,708	15,708
R^2	0.008	0.091	0.019	0.102	0.019	0.102

Notes. The table reports regressions based on Equation (3), where the dependent variable is an indicator variable taking the value one if the taxpayer self-reports any foreign income for the follow-up year. In columns (1) and (2), the estimated effects are relative to the pooled sample of all treatment groups; in columns (3)–(6), the estimated effects are relative to the *Base* treatment. The indicator variables *No letter*, *Fairness*, *Societal benefits*, and *Detection* take the value one if the taxpayer is in the respective treatment. The indicator variable *Moral* takes the value one if the taxpayer is in the *Fairness* treatment or the *Societal benefits* treatment. The reported *p*-value from *F*-test is for the hypothesis that all *Moral* treatments have the same effect. Columns (2), (4), and (6) include the following controls: an indicator variable taking the value one if the taxpayer self-reported any foreign income for the baseline year, the amount of foreign income recorded in the ACA-reports for 2011 and 2012 (scaled in units of one million NOK), gender, age (an indicator variable taking the value one if the taxpayer is more than 60 years), an indicator variable taking the value one if the taxpayer is a Norwegian citizen, and an indicator variable of socioeconomic status taking the value one if the taxpayer is in the upper 25% of the income and wealth distribution at baseline. SES, socioeconomic status.

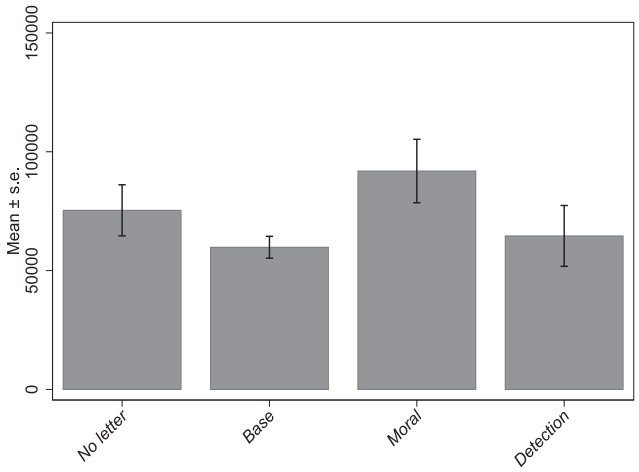
Robust standard errors are in parentheses: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In columns (3)–(6), we observe that the detection letter has a large and statistically significant effect on the extensive margin ($p < 0.001$, column (6)), increasing the share of individuals who self-report foreign income from 20% and 33%. In contrast, the *Moral* treatments do not on average have a significant effect on the extensive margin ($p = 0.122$, column (6)). In panel (c) of Figure 2, we show that these patterns are robust to only considering the effect on the taxpayers that self-report some foreign income but do not request any deductions.

Result 3. *The moral letters and the detection letter have very different effects on the extensive margin. The detection letter causes a large and statistically significant increase in the share of individuals who self-report foreign income, whereas the moral letters on average have no effect on the extensive margin.*

In Figure 3, we show the average self-reported foreign income for the group that reports a positive amount in each treatment. We observe that the conditional average foreign income reported is about 50% higher in the *Moral* treatments than in the *Base* treatment, whereas we find no difference between the *Base* treatment and the *Detection* treatment. In fact, the conditional average foreign income reported is lower in the *Base* treatment and the *Detection* treatment than in the *No letter* group. When interpreting these differences, however, it is important to keep in

Figure 3. Self-Reported Foreign Income, Conditional on It Being Positive



Note: The figure shows the average amount of self-reported foreign income in NOK for the subset of taxpayers that report a positive foreign income, by treatment.

mind how the treatments affected the extensive margin, as reported in Result 3. The moral treatments did not have a significant effect on the extensive margin, which means that the conditional average foreign income captures the effect on the intensive margin. Given the assumptions outlined in Section 5.2 and the estimates in Table 6 (column 6), it follows that the estimated intensive margin effect for the moral treatments is 37,900 NOK (with a standard error of 14,300 NOK). For the *Detection* treatment, however, it is not possible to provide a point estimate of the effect on the intensive margin with our data, since this treatment also had a large effect on the extensive margin. However, using the decomposition of the average treatment effect in Section 5.2, it follows that the estimated intensive margin effect for the *Detection* treatment is between 0 NOK and 53,400 NOK (with a standard error of 24,300 NOK on the upper bound).

In Tables 6 and 7, we observe that the inclusion of the background variables does not change the estimated treatment effects, but some of the background variables are predictive for how much is self-reported for the follow-up year. We observe that there is a highly significant positive association between self-reported foreign income for the baseline year and for the follow-up year: taxpayers who self-report foreign income for the baseline year are more likely to self-report foreign income for the follow-up year and the self-reported amount is increasing in the amount they self-reported for the baseline year. We also observe that the foreign income recorded in the ACA-reports is positively associated with self-reported foreign income, but we only find a statistically significant relationship at the extensive margin. This may reflect that the level of foreign income in the ACA-reports is a noisy measure of actual foreign income, since these reports only contain information from some countries and may even for these countries have an imprecise measure of the taxpayer's actual income. Interestingly, conditional on the ACA-reports, we find that females and older people tend to report higher levels of foreign income and are also more likely to report foreign income, which is consistent with the common finding that these personal characteristics are positively associated with being morally motivated (Andreoni et al. 1998, Cappelen et al. 2015). We also observe that individuals with higher income or greater wealth self-report higher foreign income, whereas we do not find that Norwegian citizens are significantly different in their self-reporting behavior than other citizens.

6.2. Heterogeneity Analysis

In Table 8, we report estimated treatment effects by subgroup based on Equation (4), where we focus on the *Moral* treatments combined (panel A) and the *Detection* treatment (panel B) compared with the *Base* treatment. The most striking feature of this analysis is the consistency in the estimated treatment effects: for all subgroups, we observe that the moral letters and the detection letter increase the level of self-reported foreign income. Not surprisingly, since we here look at smaller subsamples, the estimated effect is not statistically significant for all subgroups, but the robust pattern speaks of these letters having increased self-reported foreign income. Similarly, for the extensive margin, we find the same patterns across subgroups as in the main analysis: the moral letters typically have a small or no effect on the extensive margin, whereas the detection letter has a large and highly statistically significant effect in all subgroups. Overall, the heterogeneity analysis clearly demonstrates that our main findings are robust across subgroups.

The fact that the patterns are similar across subgroups is also reflected in most interactions between treatments and the background characteristics not being significant, with the exception of the interaction between the *Moral* treatment and high socioeconomic status for amount reported ($p = 0.016$, panel A) and between the *Detection* treatment and being a Norwegian citizen ($p = 0.010$, panel B) or above 60 years ($p < 0.001$, panel B).

In the "*Positive base*" columns, we report the interaction between the treatment and an indicator variable for whether an individual self-reported a positive foreign income for the baseline year. In line with the finding that the *Moral* treatments primarily worked on the intensive margin, we observe in panel A that the effect of moral suasion on amount reported is particularly strong for the group that self-reported a positive amount for the baseline year ($p = 0.026$). In

Table 8. Heterogeneity in How *Moral* and *Detection* Treatments Work

Panel A: Heterogeneity: *Moral* treatments										
	Level reported (in NOK)					Reporting positive amount				
	Citizen	*Above 60*	*Female*	*High SES*	*Positive base*	*Citizen*	*Above 60*	*Female*	*High SES*	*Positive base*
Moral	9,316.5**	9,626.8	7,073.3*	634.7	2,132.3**	0.0072	0.0018	0.023**	0.0092	0.015*
	(4,717.6)	(5,905.8)	(4,141.6)	(1,629.1)	(995.1)	(0.010)	(0.0092)	(0.0097)	(0.0092)	(0.0093)
Group × *Moral*	−3,518.5	−4,059.4	668.2	14,849.2**	14,382.9*	0.0087	0.018	−0.024	0.0063	−0.0084
	(6,018.2)	(6,164.7)	(5,717.9)	(6,135.6)	(7,458.9)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Group	7,092.0	12,734.4**	7,892.6*	−6,092.3	12,412.7***	−0.0065	0.20***	0.074***	0.035***	0.16***
	(4,745.9)	(5,884.0)	(4,710.0)	(5,867.8)	(3,505.7)	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Treatment effect on *group*	5,798.1	5,567.3***	7,741.6**	15,483.9***	16,515.1**	0.016	0.020*	−0.0013	0.016	0.0070
	(3,599.7)	(1,906.5)	(3,900.3)	(5,944.7)	(7,404.4)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Observations	11,768	11,768	11,768	11,768	11,768	11,768	11,768	11,768	11,768	11,768
Panel B: Heterogeneity: *Detection* treatment										
	Level reported (in NOK)					Reporting positive amount				
	Citizen	*Above 60*	*Female*	*High SES*	*Positive base*	*Citizen*	*Above 60*	*Female*	*High SES*	*Positive base*
Detection	14,298.6	11,417.8	5,520.4	2,475.3*	6,114.7***	0.101***	0.0656***	0.135***	0.119***	0.172***
	(8,932.3)	(9,027.6)	(3,464.2)	(1,449.6)	(2,044.9)	(0.0162)	(0.0146)	(0.0154)	(0.0148)	(0.0152)
Group × *Detection*	−8,609.9	−3,405.4	9,036.8	15,654.1	7,707.6	0.0599**	0.126***	−0.00326	0.0341	−0.0918***
	(9,433.3)	(9,552.1)	(9,268.8)	(9,660.7)	(10,310.8)	(0.0233)	(0.0229)	(0.0237)	(0.0239)	(0.0239)
Group	835.5	5,082.4**	−1,840.6	4,499.8*	15,765.3***	−0.0147	0.195***	0.0718***	0.0336**	0.172***
	(2,387.8)	(2,271.3)	(1,764.9)	(2,330.6)	(2,872.1)	(0.0131)	(0.0138)	(0.0124)	(0.0132)	(0.0140)
Treatment effect on *group*	5,688.7*	8,012.4***	14,557.2*	18,129.4*	13,822.3	0.161***	0.191***	0.132***	0.153***	0.0800***
	(3,090.8)	(2,922.7)	(8,684.7)	(9,527.0)	(10,138.6)	(0.0167)	(0.0177)	(0.0180)	(0.0188)	(0.0184)
Observations	5,919	5,919	5,919	5,919	5,919	5,919	5,919	5,919	5,919	5,919

Notes. The table reports regression estimates based on Equation (4), with the same set of controls as in Table 6 and Table 7. Column headers show the indicator variable that is used to define the indicator variable *Group*, where *Group* takes the value one if the indicator variable in the heading of the respective column takes the value one: *Citizen* (the taxpayer is a Norwegian citizen), *Above 60* (the taxpayer is above 60 years old), *Female* (the taxpayer is female), *High SES* (the taxpayer is in the upper 25% of the income and wealth distribution at baseline), and *Positive base* (the taxpayer self-reported any foreign income for the baseline year). Effects are estimated relative to the *Base* treatment. Panel A reports estimates for the taxpayers who were in the *Moral* treatments or the *Base* treatment, where *Moral* is an indicator variable that takes the value one if the taxpayer is in one of the *Moral* treatments. Panel B reports estimates for the taxpayers who were in the *Detection* treatment or the *Base* treatment, where *Detection* is an indicator variable that takes the value one if the taxpayer is in the *Detection* treatment. The dependent variable is the amount of foreign income self-reported for the follow-up year (left part) and an indicator variable taking the value one if the taxpayer self-reported any foreign income for the follow-up year (right part). SES, socioeconomic status.

Robust standard errors are in parentheses: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

contrast, the *Moral* treatment had only a marginal positive effect on the taxpayers that did not report any foreign income for the baseline year, in terms of both the amount reported and the share reporting a positive amount. The pattern is strikingly different for the *Detection* treatment. In panel B, we observe that the detection letter worked on the extensive margin for both groups, but particularly for those who did not report any foreign income for the baseline year ($p < 0.001$). The *Detection* treatment also caused an increase in the amount reported in both groups, but this effect is only statistically significant for the taxpayers who did not report any foreign income for the baseline year ($p = 0.003$).

Result 4. *The effects of the moral letters and the detection letter are robust across subgroups, with few significant interaction effects between subgroups and treatment. The heterogeneity analysis provides evidence consistent with the moral letters strongly affecting the intensive margin and the detection letter strongly affecting the extensive margin of taxpayer behavior.*

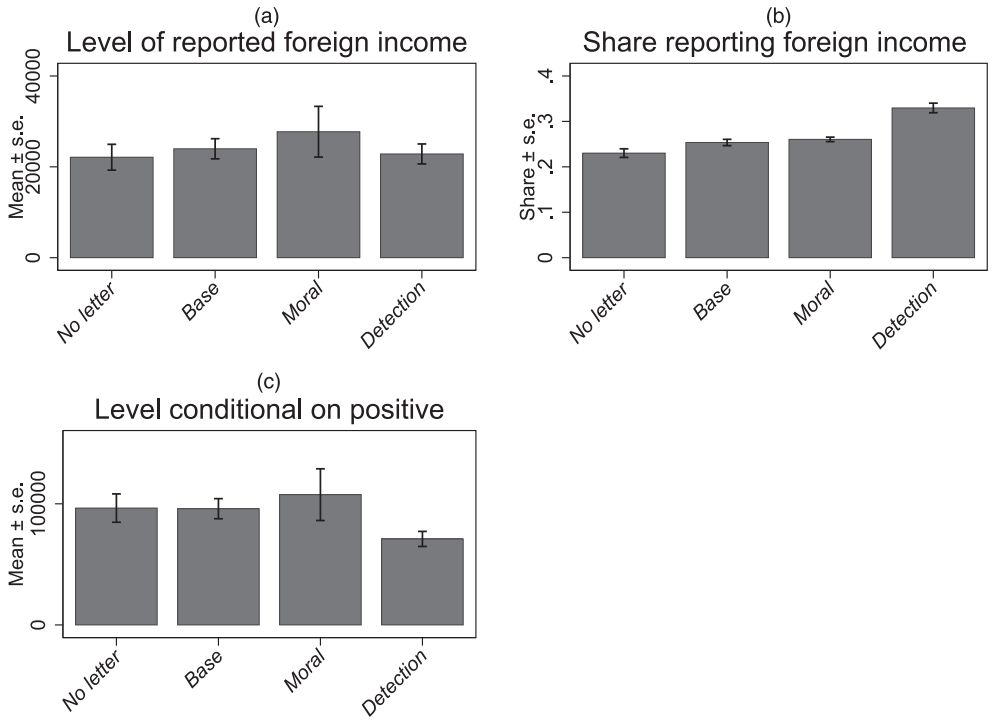
6.3. Long-Term Evidence

In this section, we study the self-reporting behavior of the taxpayers in our sample in 2014, when they had to self-report their foreign income in the prepopulated tax return for 2013. The deadline was again in week 18, which means that their choice of how much to self-report for 2013 was made more than one year after they received the treatment letters. It is therefore an open question whether the message of the letter still influences the taxpayers.

In the long run, the treatment letters may not only affect the choice of how much to self-report, but also the choice of how much income-generating activity to have abroad. However, it should be kept in mind that the taxpayers received the letter in week 15 of 2013, which means that they had limited opportunity to adjust the extent to which they earned income abroad in 2013 as a response to the intervention.

In Figure 4, we provide an overview of our long-term findings. In panel (a), we observe that the average self-reported foreign income in the *Moral* treatments is somewhat higher than in the other treatments, but this difference is not statistically significant. In panel (b), however, we observe that even in the long run, there is a large and statistically significant effect of the detection letter on the extensive margin: the share of taxpayers reporting a positive foreign income increases from 25.4% in the *Base* treatment to 32.6% in the *Detection* treatment ($p < 0.001$). We also observe that the base letter itself has some effect on the extensive margin in the long run ($p = 0.046$), but we do not find any additional effect of the moral letters ($p = 0.422$). Our findings are robust to the removal of

Figure 4. Long-Term (2013) Self-Reported Foreign Income



Note: The figure shows, for tax year 2013, the average amount of self-reported foreign income in NOK (panel (a)), the share of taxpayers who self-reported a positive foreign income (panel (b)), and the average amount of self-reported foreign income in NOK for the subset of taxpayers that reported a positive foreign income (panel (c)), by treatment.

the individuals who were most likely to be followed up by tax authorities in 2013.[2]

Interestingly, in panel (c), we observe that the average self-reported foreign income for the group that reports a positive amount is significantly lower in the *Detection* treatment than in the other treatments. This is consistent with the absence of a detection system having motivated the foreign investments in the first place and information on detection possibility making these types of investments less attractive in the long run. An alternative mechanism could be that the detection letter increases the incentive for taxpayers to seek out professional tax advisors to uncover legal ways to reduce taxable income (Slemrod et al. 2001). In line with what we should expect from panels (a) and (b), we do not observe any other significant differences in panel (c).

Overall, the long-term data provide evidence of the *Detection* treatment having a lasting effect on taxpayer behavior by significantly increasing the share of taxpayers who self-report a positive foreign income. In contrast, the *Moral* treatments primarily seem to have an effect in the short term.

Result 5. *The detection letter has a significant long-term effect on the extensive margin, whereas we do not find any significant long-term effects of the moral letters.*

The difference in long-term effects of the detection letter and the moral letters may speak to the underlying mechanisms driving the initial effects observed for the follow-up year. The *Moral* treatments may primarily have made moral arguments salient when the taxpayers were due to report for the follow-up year, without causing a fundamental change in the preferences of the individual and therefore not causing a change in their long-term behavior. The *Detection* treatment, on the other hand, may have caused the taxpayers to update their beliefs about the detection probability, and these updated beliefs may have been sustained also in the long run.

7. Conclusions

Our study shows that tax administrations should consider a rich set of instruments in the fight against tax evasion. A simple and cheap field intervention using letters increased the amount of self-reported foreign income by around 122 million NOK (approximately 21 million USD) for the follow-up year. The intervention also cleanly identified that both moral motivation and economic incentives play a role in shaping taxpayer behavior. In line with the increasing focus among tax administrators on building a tax morale in society (Luttmer and Singhal 2014), we find a large effect of moral suasion, of the same size as the effect of including a sentence that is likely to have increased the perceived probability of detection. However, moral appeals and detection probability influence tax behavior in different ways. The moral appeals only affect the intensive margin, whereas increasing the detection probability strongly affects the extensive margin. We also report long-term effects of the intervention, where we show that the detection letter has a large effect on the extensive margin even one year after the taxpayer received the letter, whereas the moral letters have no statistically significant long-term effects.

The long-term findings show that it is important to distinguish between (at least) two different mechanisms when considering how moral suasion may reduce tax evasion. In our study, it appears that moral suasion mainly worked by making the moral argument salient when the taxpayer made the decision of how much to report, but did not work at a more fundamental level by increasing the weight taxpayers assign to the moral cost of misreporting (since there was no effect of the moral letters in the long run). Our study also demonstrates that detection probability plays a critical role for tax compliance, and an important avenue for future research would be to study how the moral motive and the detection motive interact in shaping moral behavior. Is there crowding out of moral motivation among taxpayers when tax administrations primarily focus on detection probability and penalty rates (Gneezy and Rustichini 2000)? The answer to this question might, as suggested by Kirchler et al. (2008), depend on the tax climate in a society. Moreover, the fact that the detection letter and the moral letters worked at different margins shows that the context is important when considering different strategies for increasing tax compliance. A focus on tax morale will only work when taxpayers consider it morally wrong to cheat on taxes, but may then have significant impact by making taxpayers report income when it is difficult for the tax administration to detect tax evasion. A focus on detection probability is likely to increase tax compliance also in settings where tax subjects are not morally motivated, but may cause a more narrow response where taxpayers only report income they believe is likely to be detected by the tax administration.

Our preferred interpretation of the treatment effect of the moral letters is that these letters made the moral argument more salient and thereby affected the behavior of the taxpayers, but we cannot rule out that the effect of the moral letters go through other channels. It is, for example, possible that the moral letters affected the taxpayer's perception of the tax administration and that this in turn affected their willingness to comply (Feld and Frey 2007). Another possibility is that the moral letters had an effect by changing the mood of the taxpayers, which would be in line with the finding from laboratory experiments that emotions affect prosocial behavior (Drouvelis and

Grosskopf 2016). An important challenge for future research is therefore to provide further evidence on how moral appeals shape human behavior.

Finally, the study contributes to the broader discussion in economics about the importance of moral motivation. A growing literature in behavioral economics, mostly relying on laboratory experiments, has documented that moral motivation matters for people when making economic decisions (Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Konow 2000; Andreoni and Miller 2002; Charness and Rabin 2002; Engelmann and Strobel 2004; Cappelen et al. 2007, 2013; Fehr et al. 2013). Moreover, in a related literature it has been shown that people do not always lie even when they have the opportunity to do so and can gain from it (Gneezy 2005, Erat and Gneezy 2012). These laboratory experimental findings have sometimes been challenged, because they are established in an artificial setting with small stakes (Levitt and List 2007). We demonstrate that moral motivation not only matters in the laboratory, but also in field settings involving large stakes. A simple moral message caused the taxpayers to self-report a significantly larger amount of foreign income, which illustrates the power of moral motivation in shaping human behavior. This insight is clearly important for policymakers considering how to reduce tax avoidance, but also more broadly. Our research suggests that moral appeals should be part of the toolkit of any organization in which compliance is an important challenge.

Acknowledgments

The authors thank Nadja Dwenger, Uri Gneezy, Espen Moen, David Laibson, Camille Landais, John List, Ragan Petrie, Agnar Sandmo, László Sándor, Emmanuel Saez, Guttorm Schjelderup, Paul Smeets, and Kjetil Storesletten for extremely valuable comments and suggestions; Kenneth Andreassen, Ingvild Holler Deisz, Marta Johanne Gjengedal, Anita Hallenstvedt, Pia Høst, Elin Imsland, Paul Gunnar Larssen, Dag Løvås, Bente Lundekvam, Sissel Madsen, Gitte Mårtensen, Vibeke Vik Nordang, Elin Aam Svendsen, Irene Søreide, Thomas Tangen, Christine Osen Tefre, Marcus Zackrisson, and others at the Norwegian Tax Authority for their assistance in implementing the experiment; and Trine Sivertsen Sommerlade at Bergen Brannvesen and Fredrik Naumann for pictures used in the experimental materials.

Endnotes

[1] These results are robust to the different versions of the *Societal benefits* treatments (verbal, visual, verbal and visual). In all cases, we observe an increase in self-reported foreign income, and we cannot reject that the three versions have the same effect on taxpayer behavior ($p = 0.59$).

[2] This robustness analysis was based on communication with NTA, who provided us with information about their auditing rules. Note that according to our agreement with NTA, no taxpayer was followed up with before they had submitted their tax return for the follow-up year. Hence, auditing from NTA could only potentially affect our long-term findings.

References

Allingham MG, Sandmo A (1972) Income tax evasion: A theoretical analysis. *J. Public Econom.* 1(3–4):323–338.

Alm J, Torgler B (2006) Culture differences and tax morale in the United States and in Europe. *J. Econom. Psych.* 27(2):224–246.

Alm J, McClelland GH, Schulze WD (1992) Why do people pay taxes? *J. Public Econom.* 48(1):21–38.

Alm J, Sanchez I, de Juan A (1995) Economic and noneconomic factors in tax compliance. *Kyklos* 48(1):3–18.

Andreoni J, Miller J (2002) Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70(2):737–753.

Andreoni J, Erard B, Feinstein J (1998) Tax compliance. *J. Econom. Literature* 36(2):818–860.

Ariel B (2012) Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology* 50(1):27–69.

Artavanis N, Morse A, Tsoutsoura M (2016) Measuring income tax evasion using bank credit: Evidence from Greece. *Quart. J. Econom.* 131(2):739–798.

Bartke S, Friedl A, Gelhaar F, Reh L (2017) Social comparison nudges—Guessing the norm increases charitable giving. *Econom. Lett.* 152:73–75.

Blumenthal M, Christian C, Slemrod J (2001) Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax J.* 54(1):125–138.

Dal Bó E, Dal Bó P (2009) "Do the right thing:" The effects of moral suasion on cooperation. *J. Public Econom.* 117(1):28–38.

Bobek DD, Hatfield RC (2003) An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. *Behav. Res. Accounting* 15(1):13–38.

Bolton GE, Ockenfels A (2000) ERC: A theory of equity, reciprocity, and competition. *Amer. Econom. Rev.* 90(1):166–193.

Bryan CJ, Adams GS, Monin B (2012) When cheating would make you a cheater: Implicating the self prevents unethical behaviour. *J. Experiment. Psych. General* 142(4):1001–1005.

Cappelen AW, Hole AD, Sørensen EØ, Tungodden B (2007) The pluralism of fairness ideals: An experimental approach. *Amer. Econom. Rev.* 97(3):818–827.

Cappelen AW, Moene KO, Sørensen EØ, Tungodden B (2013) Needs versus entitlements: An international fairness experiment. *J. Eur. Econom. Assoc.* 11(3):574–598.

Cappelen AW, Nygaard K, Sørensen EØ, Tungodden B (2015) Social preferences in the lab: A comparison of students and a representative population. *Scandinavian J. Econom.* 117(4):1306–1326.

Castro L, Scartascini C (2015) Tax compliance and enforcement in the Pampas—Evidence from a field experiment. *J. Econom. Behav. Organ.* 116(1):65–82.

Cebula RJ, Feige EL (2012) America's unreported economy: Measuring the size, growth and determinants of income tax evasion in the U.S. *Crime Law Soc. Change* 57(3):265–285.

Charness G, Rabin M (2002) Understanding social preferences with simple tests. *Quart. J. Econom.* 117(3):817–869.

Chetty R, Looney A, Kroft K (2009) Salience and taxation: Theory and evidence. *Amer. Econom. Rev.* 99(4):1145–1177.

Cobham A, Jansky P (2018) Global distribution of revenue loss from corporate tax avoidance: re-estimation and country results. *J. Internat. Development* 30(2):206–232.

Coleman S (1996) The Minnesota income tax compliance experiment: State tax results. Munich Personal RePEc Archive Paper 4827, University Library of Munich, Germany.

Coleman S (2007) The Minnesota income tax compliance experiment: Replication of the social norms experiment. Munich Personal RePEc Archive Paper 5820, University Library of Munich, Germany.

Cullis JG, Lewis A (1997) Why people pay taxes: From a conventional economic model to a model of social convention. *J. Econom. Psych.* 18(2–3):305–321.

Del Carpio L (2013) Are the neighbours cheating? Evidence from a social norm experiment on property taxes in Peru. Working paper, INSEAD, Fontainebleau, France.

Drouvelis M, Grosskopf B (2016) The effects of induced emotions on pro-social behaviour. *J. Public Econom.* 134(1):1–8.

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

Dwenger N, Kleven H, Rasul I, Rincke J (2016) Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *Amer. Econom. J. Econom. Policy* 8(3): 203–232.

Engelmann D, Strobel M (2004) Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *Amer. Econom. Rev.* 94(4):857–869.

Engström P, Holmlund B (2009) Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Appl. Econom.* 41(19): 2419–2430.

Erat S, Gneezy U (2012) White lies. *Management Sci.* 58(4):723–733.

Fehr E, Gächter S (1998) Reciprocity and economics: The economic implications of Homo Reciprocans. *Eur. Econom. Rev.* 42(3–5): 845–859.

Fehr E, Schmidt KM (1999) A theory of fairness, competition and cooperation. *Quart. J. Econom.* 114(3):817–868.

Fehr E, Glätzle-Rützler D, Sutter M (2013) The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *Eur. Econom. Rev.* 64(1):369–383.

Feld LP, Frey BS (2007) Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law Policy* 29(1):102–120.

Fellner G, Sausgruber R, Traxler C (2013) Testing enforcement strategies in the field: Threat, moral appeals and social information. *J. Eur. Econom. Assoc.* 11(3):634–660.

Foss C, Thorsager M, Olsen Ø (2015) Ny måte å kontrollere selvangivelsen på—fra fokus på enkeltposter til skattytere som gjør feil [in Norwegian]. *Skatteetatens Analysenytt* 5(1):6–9.

Gneezy U (2005) Deception: The role of consequences. *Amer. Econom. Rev.* 95(1):384–394.

Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. *Quart. J. Econom.* 115(3):791–810.

Graetz MJ, Reinganum JF, Wilde LL (1986) The tax compliance game: Toward an interactive theory of law enforcement. *J. Law Econom. Organ.* 2(1):1–32.

Hallsworth M (2014) The use of field experiments to increase tax compliance. *Oxford Rev. Econom. Policy* 30(4):658–679.

Hallsworth M, List JA, Metcalf RD, Vlaev I (2017) The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *J. Public Econom.* 148(1):14–31.

Kirchler E (2007) *The Economic Psychology of Tax Behaviour* (Cambridge University Press, Cambridge, UK).

Kirchler E, Hoelzl E, Wahl I (2008) Enforced versus voluntary tax compliance: The "slippery slope" framework. *J. Econom. Psych.* 29(2):210–225.

Kleven HJ, Knudsen MB, Kreiner CT, Pedersen S, Saez E (2011) Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica* 79(3):651–692.

Konow J (2000) Fair shares: Accountability and cognitive dissonance in allocation decisions. *Amer. Econom. Rev.* 90(4):1072–1091.

Krupka E, Weber R (2009) The focusing and informational effects of norms on pro-social behavior. *J. Econom. Psych.* 30(3):307–320.

Lamberton C, De Neve J-E, Norton MI (2018) The power of voice in stimulating morality: Eliciting taxpayer preferences increases tax compliance. *J. Consumer Psych.* 28(2):310–328.

Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspectives* 21(2):153–174.

Luttmer, EP, Singhal M (2014) Tax morale. *J. Econom. Perspectives* 28(4):149–168.

Manski CF (1997) Monotone treatment response. *Econometrica* 65(6): 1311–1334.

Mascagni G (2018) From the lab to the field: A review of tax experiments. *J. Econom. Surveys* 32(2):273–301.

Pomeranz D (2015) No taxation without information: Deterrence and self-enforcement in the value added tax. *Amer. Econom. Rev.* 105(8):2539–2569.

Sandmo A (2005) The theory of tax evasion: A retrospective view. *National Tax J.* 58(4):643–663.

Sandmo A (2012) An evasive topic: Theorizing about the hidden economy. *Internat. Tax Public Finance* 19(1):5–24.

Slemrod J (2007) Cheating ourselves: The economics of tax evasion. *J. Econom. Perspectives* 21(1):25–48.

Slemrod J, Yitzhaki S (2002) Tax avoidance, evasion, and administration. Auerbach AJ, Feldstein M, eds. *Handbook of Public Economics*, vol. 3 (Elsevier, Amsterdam), 1423–1470.

Slemrod J, Blumenthal M, Christian C (2001) Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *J. Public Econom.* 79(3):455–483.

Staub KE (2014) A causal interpretation of extensive and intensive margin effects in generalized Tobit models. *Rev. Econom. Statist.* 96(2):371–375.

Torgler B (2002) Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *J. Econom. Surveys* 16(5):657–683.

Torgler B (2004) Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Econom. Governance* 5(3):235–253.

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

Wenzel M (2004) An analysis of norm processes in tax compliance. *J. Econom. Psych.* 25(2):213–228.