Using behavioral insights and experimentation to prevent APP fraud

Preliminary analysis

February 2021

1 Introduction

This document presents preliminary findings from two experiments that investigate how to reduce the likelihood that people fall for APP fraud.

Both experiments involved inviting individuals to take part in an online survey. Upon entering the survey, participants answered a series of questions about themselves and their use of online banking. Participants were then provided with information about three hypothetical payments and were offered the opportunity to either make the payments or to cancel them. They could make these payments via an app that was accessible within the survey. Further, participants were incentivised to behave as they would in the real world; they earned more if they made 'legitimate' payments and earned less if they made 'fraudulent' payments.

In the first experiment, participants were able to make these payments using a mobile banking app. Participants were randomly allocated to groups that were shown different versions of the banking app.

In the second experiment, participants were able to make these payments using a Payment Initiation Service Provider (PISP) app. Participants were randomly allocated to groups that were shown different versions of the PISP app.

The remainder of this document is structured as follows. Sections 2 and 4 describe the experimental designs. Sections 3 and 5 present the results from the experiments. Section 6 concludes by summarising the main findings.

2 Experiment 1: Methodology

The study begins by asking participants questions about: their age, their gender, which region they live in, their income, their use of online banking, and which banks they use. Participants are then presented with the following instructions:

> **You will have the opportunity to make three payments using a mobile banking app.**
>
> We will present you with descriptions of the payments (scenarios) and will then take you to an app where you can decide to make (or cancel) the payments. Some of the payments might not be legitimate.
>
> The amount that you can earn while taking this survey depends on your decisions:
> 1. If you **make** a payment that is **legitimate**, you will earn 50% of the amount paid.
> 2. If you **make** a payment that is **not legitimate**, you will lose 100% of the amount paid.
> 3. If you **cancel** a payment that is **legitimate**, you will lose the sum of money listed in the payment description.
> 4. If you **cancel** a payment that is **not legitimate**, you lose/gain nothing.
>
> **The amount that you earn while taking this survey will be converted into actual money**, which will be paid to you at a rate of £1,000 (in the survey) = £1 (in real life). You cannot earn less than £0 when participating in this survey.

We then measured participants' understanding of the instructions using five comprehension questions. Finally, they were told that:

> To make a payment, you will have to use a mobile bank app that will be presented to you.
>
> Not all buttons and features in the app will be clickable, and the information from the payment scenario will have been pre-filled. If you cancel the payment while using the app, you will be taken to the next scenario (there are **three** scenarios).
>
> **Click next to begin reading the first scenario.**

We then proceeded to provide participants with descriptions of the first payment scenario.

All participants were asked to read through three payment scenarios, each representing a different category of payment:

1. Paying an invoice for the remodelling of your kitchen
2. Buying a laptop from a seller on the Facebook Marketplace
3. Paying overdue self-declared taxes to HMRC

We randomised the order in which payments from these categories appeared. We developed one fraudulent and one legitimate payment scenario for each category, and participants were only shown one fraudulent scenario. We randomised whether participants were shown a fraudulent version of scenario 1 (invoice), 2 (laptop), or 3 (HMRC).

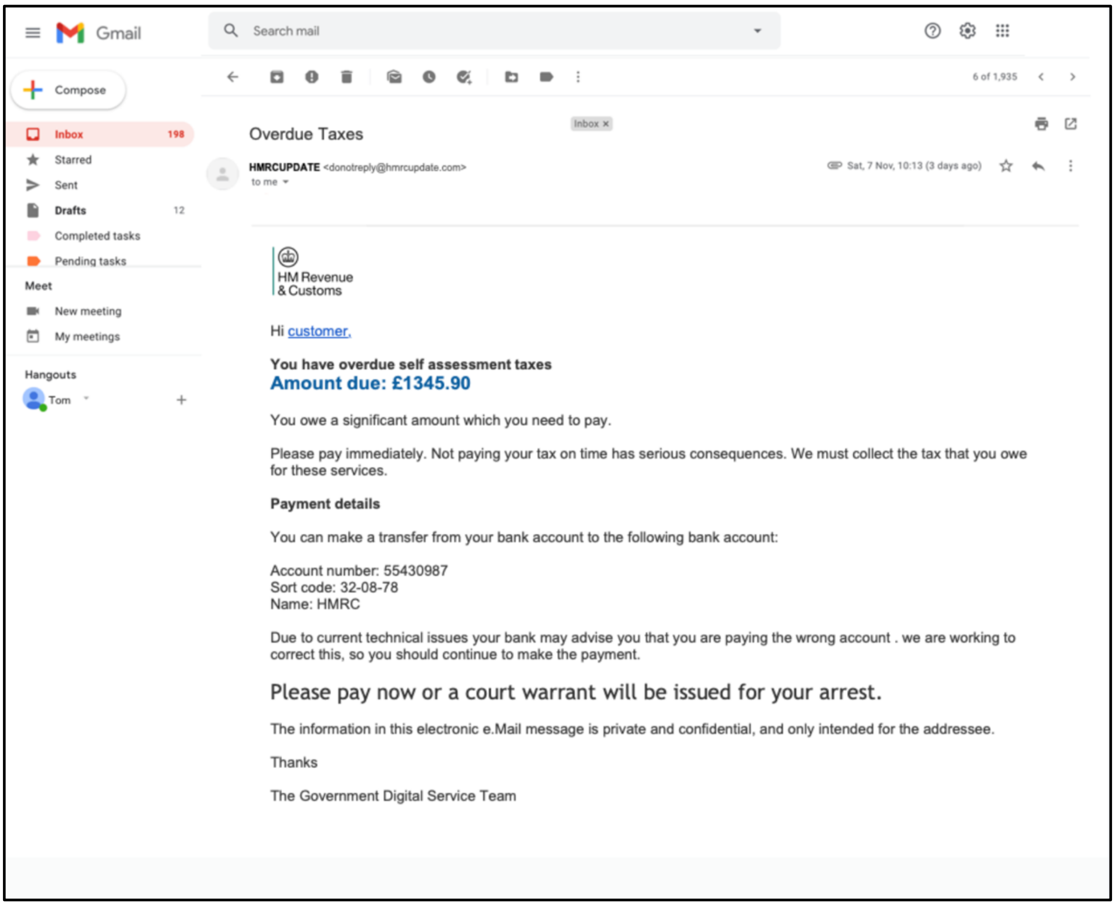
Below we present an example of what the fraudulent HMRC payment scenario looked like:

You usually file Self-Assessment tax returns with HMRC but have not yet paid your tax this year.

You just received an email explaining that you have not paid the balance due (£1345.90), with instructions regarding how to pay. The amount mentioned in the email is roughly what you expected to have to pay in tax this year.

Please review the email and proceed with the payment as you see fit.

Making the payment will earn you 50% of the value of the payment. If you do not pay, you may incur a loss of £5,000 (from costs associated with paying your taxes late).



Overdue Taxes

HMRCUPDATE <donotreply@hmrcupdate.com>
to me

Sat, 7 Nov, 10:13 (3 days ago)

HM Revenue
& Customs

Hi customer,

You have overdue self assessment taxes
Amount due: £1345.90

You owe a significant amount which you need to pay.

Please pay immediately. Not paying your tax on time has serious consequences. We must collect the tax that you owe for these services.

Payment details

You can make a transfer from your bank account to the following bank account:

Account number: 55430987
Sort code: 32-08-78
Name: HMRC

Due to current technical issues your bank may advise you that you are paying the wrong account . we are working to correct this, so you should continue to make the payment.

Please pay now or a court warrant will be issued for your arrest.

The information in this electronic e.Mail message is private and confidential, and only intended for the addressee.

Thanks

The Government Digital Service Team

After reading a given scenario description and viewing the attached evidence (e.g., email receipts, invoices), participants were given the opportunity to make the payment using a mobile banking app. Participants were randomly allocated to groups that were shown different versions of the banking app. Any given participant was shown the same app when making all three payments.

The banking app that participants were shown was inspired by a recent version of the Santander banking app. Participants who were allocated to the 'control' group were shown a version of the banking app that stayed true to Santander's app. Other participants were allocated to groups that were shown apps that differed in the following way:

1. Whether the app uses a risk-based approach (involves eliciting more information about the payment and only presenting warnings when a payment is deemed to be risky).
2. Whether the app has amended Calls to Action (CTAs); this involves including more ‘buttons’ within the app that let participants cancel a payment, save the payment for later, or make a call to the bank.
3. Whether the app includes ‘behavioural interventions’ (e.g., the app includes text that leverages loss aversion by stating that they might *lose* a given amount of money if they proceed).

In total, participants could be allocated to eight different app groups:

1. Control group
2. Control group + behavioural interventions
3. Control group + CTAs
4. Control groups + behavioural interventions + CTAs
5. Risk-based group
6. Risk-based group + behavioural interventions
7. Risk-based group + CTAs
8. Risk-based groups + behavioural interventions + CTAs

Further, the Risk groups were split in two: 1) high accuracy, meaning that the risk-based approach did not misclassify high-risk scenarios as low-risk or vice versa and 2) low accuracy, meaning that the risk-based approached produced some false positives (i.e., classified legitimate scenarios as being risky).

The apps were semi-interactive (i.e., participants could click most buttons and features within the apps but could not enter text or numbers as these were pre-filled).

This experimental design allowed us to isolate the average efficacy app variants, while also measuring whether there are interactive effects (e.g., whether CTAs work better in the presence of behavioural interventions and vice versa).

Participants were asked to answer a number of survey questions once they had made decisions regarding the three payment scenarios.

Our main outcome of interest is whether participants completed fraudulent and legitimate payments. Secondary outcomes of interest include whether participants liked the app, and the time it took to complete a payment.

We are recruiting 10,000 participants for the experiment. The sample is nationally representative of the adult UK population, and was recruited via Panelbase and Prolific Academic. We had not yet recruited the full sample at the time of writing (we were still missing around 1,400 individuals).

3 Experiment 1: Results

3.1 Main analysis

We begin by examining the effects of the treatments on the share of participants that made fraudulent and legitimate payments in each experimental group. As is shown in Table 1, around 22% of participants made a fraudulent payment in the control group. Those in the control group also, on average, completed 57% of legitimate payments (as there were two legitimate payments available, this means that these participants on average made 1.14 legitimate payments).

Table 1. Treatment effects on payment behaviour

	(1) Share that made a fraudulent payment	(2) Average share of legitimate payments made
Control + behavioural	-0.04**	-0.06***
	(0.02)	(0.02)
Control + CTA	-0.12***	-0.14***
	(0.02)	(0.02)
Control + behavioural + CTA	-0.14***	-0.20***
	(0.01)	(0.02)
Risk-based	0.01	0.08***
	(0.02)	(0.02)
Risk-based + behavioural	0.01	0.10***
	(0.02)	(0.02)
Risk-based + CTA	-0.18***	-0.07***
	(0.01)	(0.02)
Risk-based + behavioural + CTA	-0.16***	-0.08***
	(0.01)	(0.02)
Average in the control group	0.22***	0.57***
	(0.01)	(0.01)
Observations	8606	8606
R-squared	0.047	0.059

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are (1) whether participants made a fraudulent payment (a binary variable that can take the values 0 and 1), and (2) the share of legitimate payments made per participant (semi-continuous variable that can take the values 0, 0.5, or 1). We obtain the share that fell for fraud in the experimental groups by adding the average in the control group to the respective coefficients.

We find that five of the seven treatments had a statistically significant effect on the share that made a fraudulent payment. The ‘Risk-based + CTA’ group had the largest effect—an 18-percentage-point decrease relative to the control group. This means that only 4% of participants in this group fell for fraud. The second-most successful intervention was the ‘Risk-based + behavioural + CTA’ group (with an effect of 16 percentage points) and the third-most successful group was the ‘Control + behavioural + CTA’ group (with an effect of 14 percentage points). We can thus conclude that the ‘CTA’ element seems to have a major impact on the share that fall for fraud, regardless of whether it is placed in a context with a risk-based approach or in an app with behavioural interventions.

While the CTA interventions generated large reductions in the share that fall for fraud, they also had an unintended side effect: they reduced the share of legitimate payments that individuals complete. For example, the CTA intervention reduced the share of legitimate payments made by 14 percentage points when applied to the control group, and by 15 points when applied to the risk-based group.

The ‘behavioural’ text reduced the share that fell for fraud by 4 percentage points when applied to the control group but had no effect when applied to the risk-based group. Further, the ‘Risk-based’ group did not have a significant effect on the share that fell for fraud. However, both the ‘Risk-based’ and ‘Risk-based + behavioural’ groups significantly *increased* the share of legitimate payments that participants completed by 8 and 10 percentage points, respectively.

We thus face a trade-off when deciding which variants to recommend. The only groups that do not induce ‘unintended consequences’ are the ‘risk-based’ and ‘risk-based + behavioural’ groups, as they increase the share of legitimate payments made. However, these groups do not have an effect on fraud. That said, while the risk-based approach might not be suitable to implement on its own, it significantly reduces the potential ‘side effects’ associated with the CTA approach. Further, it does this without negatively affecting the extent to which the CTA intervention combats fraud.

The inability of the risk-based approach to *reduce* fraud on its own may suggest that individuals do not necessarily need more information warning them about the prevalence of, or risks associated with, fraud. Instead, they need to be provided with well-placed calls to action that remind them that cancelling a payment is an option, and which provide them with an easy way of doing so if they feel suspicious.

An alternative interpretation is that—relative to the control group—the risk-based approach primarily involved removing the ‘standard’ warnings when they were not warranted, which explains the positive effect on legitimate payments, while the warnings that were triggered during ‘risky’ payments were insufficiently convincing.

It is worth noting that we would not necessarily record the same treatment effects in the real world. Some of the ‘legitimate’ scenarios in the experiment may have seemed suspicious, and we may thus be *overestimating* the negative effects of the treatments on the share that completed legitimate journeys.

3.2 Secondary analysis

We conducted a number of secondary analyses that provide us with a more nuanced understanding of the effects of the treatments and of how they operated. For example, we do not find that varying the accuracy of the risk-based approach has a significant effect on the share of fraudulent journeys that participants completed. However, we did find that the accuracy of the risk-based approach influenced the likelihood that people completed legitimate journeys (i.e., if it was less accurate, they completed around 3 percentage points fewer legitimate journeys).

We do not find that the treatments had any effects on the amount of time that it took for participants to complete the payments. Further, we find that 72 percent of users said they would prefer the ‘Risk + behavioural + CTA’ app to their existing bank app

(68% said the same in the control group). The ‘Risk + behavioural + CTA’ version of the bank app also scored the best on other customer satisfaction and usability metrics.

Across all groups, we found that people were most likely to fall for the fraudulent laptop scenario, and second-most likely to fall for the fraudulent HMRC scenario. Participants were also more likely to fall for fraud if fraud occurred when participants were completing the first scenario. Participants were least likely to fall for fraud if fraud appeared in the third scenario.

We conducted an interactivity analysis to understand if the efficacy of the treatments depended on whether participants were exposed to fraud in the first, second, or third scenario. The motivation for conducting this analysis that individuals may be ‘desensitised’ to the warnings conveyed in the app after having gone through two scenarios, making the warnings less effective if fraud appears in the third scenario. The only interactive effect that we find is that the ‘behavioural’ treatments work less well if fraud appears in the third scenario.

We present the effects of the treatments on the likelihood that participants fell for fraud if they were shown a fraudulent version of the first, second, and third scenarios, respectively, in Table 2. As we can see, there seem to be some differences in treatment efficacy by scenario. For example, the ‘Risk-based + CTA’ group is particularly effective for scenario 2 (the laptop purchase) with an effect size of 26 percentage points.

Table 2. Treatment effects by scenario

	(1) Share that made a fraudulent payment (scenario 1)	(2) Share that made a fraudulent payment (scenario 2)	(3) Share that made a fraudulent payment (scenario 3)
Control + behavioural	-0.05*	-0.06*	-0.01
	(0.03)	(0.03)	(0.03)
Control + CTA	-0.10***	-0.16***	-0.11***
	(0.02)	(0.03)	(0.02)
Control + behavioural + CTA	-0.09***	-0.21***	-0.12***
	(0.02)	(0.03)	(0.02)
Risk-based	-0.04	0.06	0.02
	(0.03)	(0.04)	(0.03)
Risk-based + behavioural	-0.06**	0.09**	-0.01
	(0.03)	(0.04)	(0.03)
Risk-based + CTA	-0.13***	-0.26***	-0.14***
	(0.02)	(0.03)	(0.02)
Risk-based + behavioural + CTA	-0.11***	-0.26***	-0.09***
	(0.02)	(0.03)	(0.02)
Average in the control group	0.16***	0.32***	0.17***
	(0.02)	(0.02)	(0.02)
Observations	2867	2875	2864
R-squared	0.021	0.102	0.037

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are whether participants made a fraudulent payment (a binary variable that can take the values 0 and 1). Regression (1) is restricted to those who were shown a fraudulent version of scenario 1. Regression (2) is restricted to those who were shown a fraudulent version of scenario 2. Regression (3) is restricted to those who were shown a fraudulent version of scenario (3). We obtain the share that fell for fraud in the experimental groups by adding the average in the control group to the respective coefficients.

4 Experiment 2: Methodology

The second experiment is structured in the same way as the first experiment. The only difference is that participants were asked to use a PISP app instead of an online banking app when making the payments. The participants use the PISP app to initiate the payment and are then directed to their bank app to authorise the payment.

Participants could be randomised into three versions of the PISP app:

> *Variant 1:* A version where COP and CRM appear when authorising the payment using the bank app
> *Variant 2:* A version where COP appears in the PISP app, and CRM appears when authorising the payment using the bank app
> *Variant 3:* A version where COP and CRM appear in the PISP app

We are recruiting 3,000 participants for this experiment (also using a nationally representative sample of UK adults) via Prolific Academic. We have, to date, recruited around 1500 individuals.

5 Experiment 2: Results

Table 3 below mirrors table 1 but presents the effects of the three PISP variants.

Table 3. Treatment effects on payment behaviour

	(1) Share that made a fraudulent payment	(2) Average share of legitimate payments made
Variant 2	0.06**	-0.01
	(0.03)	(0.02)
Variant 3	0.01	-0.03
	(0.03)	(0.02)
Average in the Variant 1 group	0.20***	0.65***
	(0.02)	(0.01)
Observations	1478	1478
R-squared	0.004	0.001

Notes: The regressions were conducted using a Linear Probability Model (LPM). The outcomes are (1) whether participants made a fraudulent payment (a binary variable that can take the values 0 and 1), and (2) the share of legitimate payments made per participant (semi-continuous variable that can take the values 0, 0.5, or 1). We obtain the share that fell for fraud in the experimental groups by adding the average in the control group to the respective coefficients.

We find that around 20% of participants in the Variant 1 group (COP and CRM in the bank app) fell for fraud. This statistic is similar to the share that fell for fraud in the control group in experiment 1 (22%). Further, we find that individuals completed around 65% of legitimate payments in the Variant 1 group (this statistic is slightly larger than the average in the control group in experiment 1).

Variant 2 *increased* the share of individuals that fell for fraud by approximately 6 percentage points. In other words, more participants fell for fraud when the responsibility for COP was assigned to the PISP, rather than the bank. We do not find that Variant 3 (when both COP and CRM were assigned to the PISP) had a significant effect relative to Variant 1. In other words, it could be that it is detrimental to split up COP and PISP, while it does not seem to matter if COP and PISP are assigned to the PISP or the bank. Neither variant had a significant effect on the share of legitimate payments that participants completed.

Both Variant 2 and 3 performed slightly worse than Variant 1 in terms of participants' subjective usability scores (e.g., whether it was easy to cancel payments, whether the app felt safe, and whether the app felt intuitive to use). However, participants were equally likely to say that they would download and use Variant 1 as Variants 2 and 3 if given the chance.

6 Discussion

Experiment 1 shows us that small changes to bank apps have the potential to drastically reduce the share of individuals that fall for APP fraud—at least in the short-run. More specifically, it seems like biggest effects are achieved when changing the CTAs that are presented in the app, offering users more opportunities to cancel and defer payments. However, this may come at a cost (i.e., dissuading individuals from making slightly suspicious looking, albeit legitimate, payments).

Experiment 2 suggests that it does not make a big difference if the responsibility for COP and CRM are allocated to a bank or a PISP. However, it does suggest that it is detrimental to split the responsibility between the two parties. The results from experiment 2 should, however, be treated with a degree of caution as we have not yet recruited the entire sample.

We plan on conducting additional analyses for both experiments 1 and 2 once we have recruited the entire study samples. These analyses will provide us with more detail regarding how the treatments worked and whether the treatments are likely to have any unintended consequences. We will also study the generalisability of the results, conduct benefit cost analyses, and will and provide more detailed recommendations.